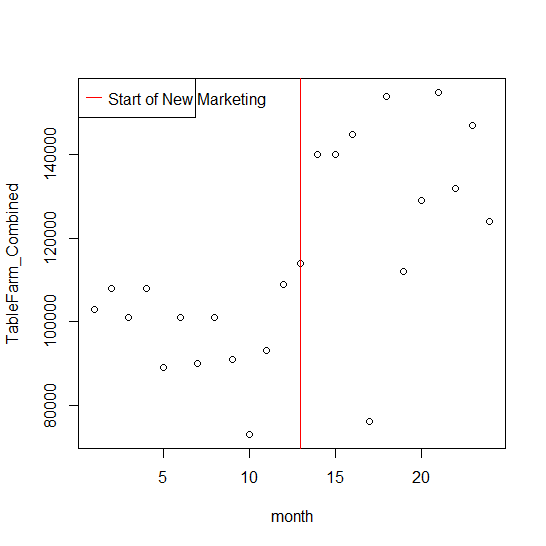
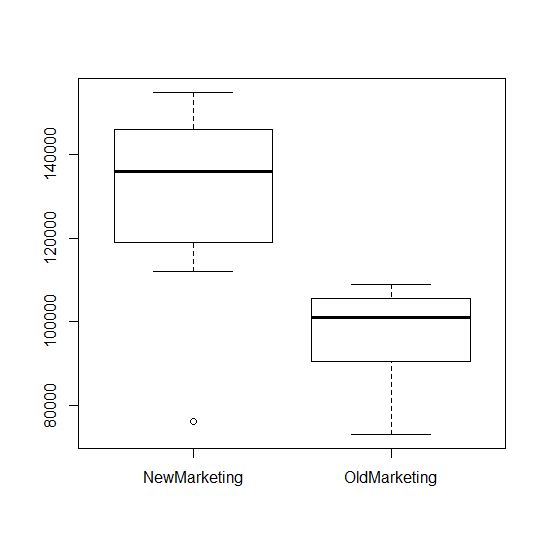
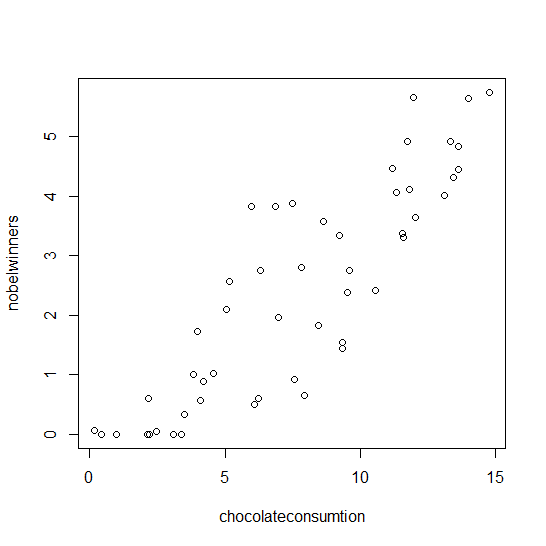
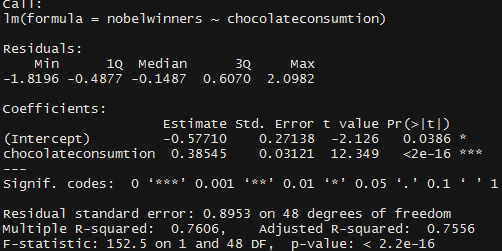
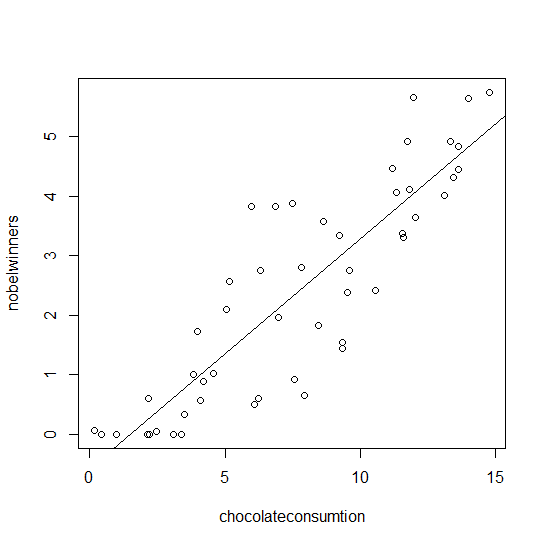
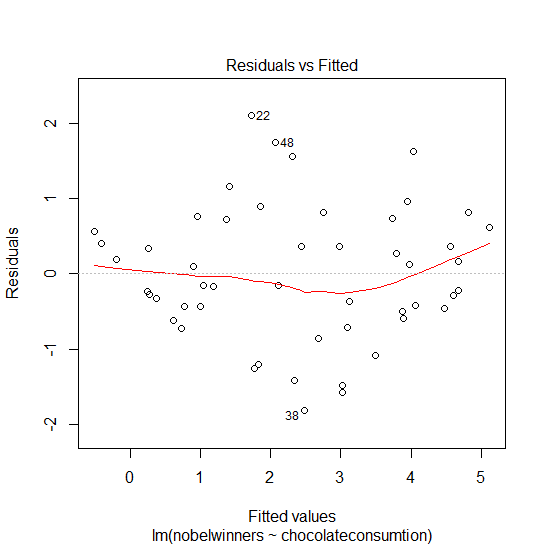
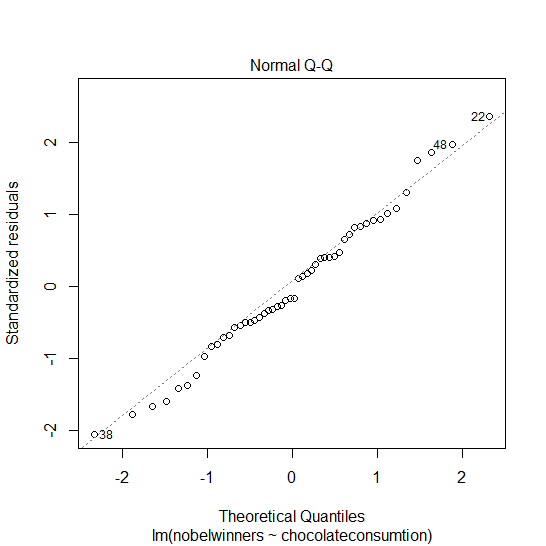
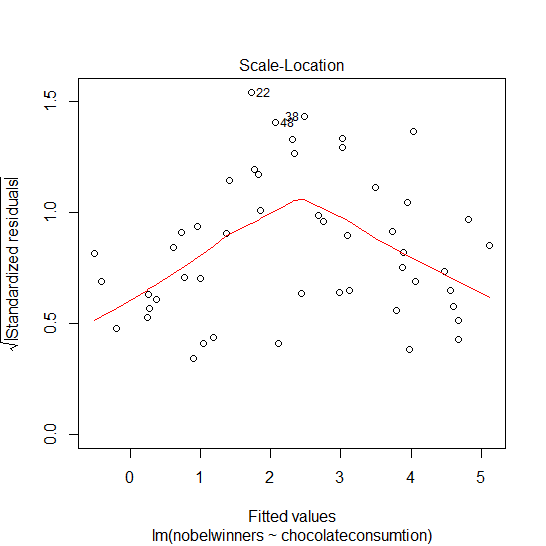
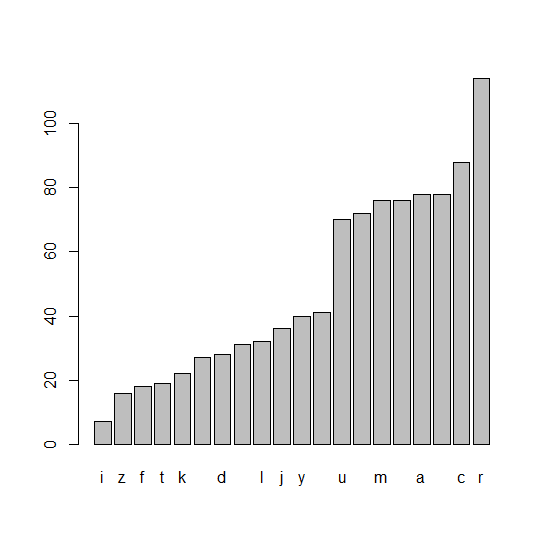
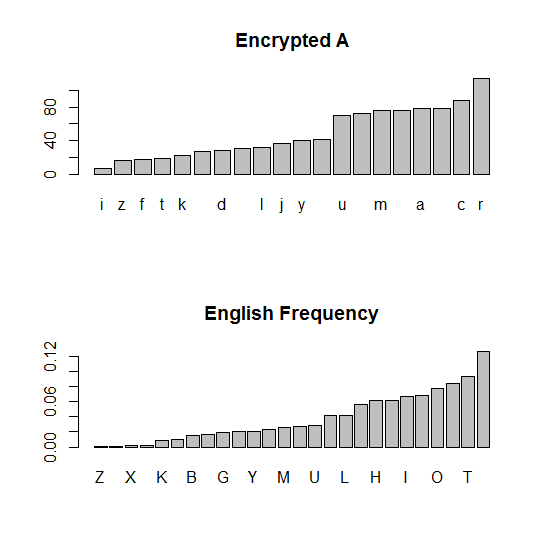
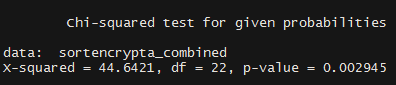
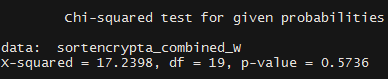
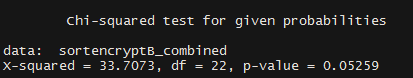
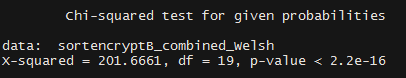
Spencer Swartz

1. Can we detect when a marketing campaign has been successful?
   * 1. TableFarm\_Combined <- c(103000,108000,101000,108000,89000,101000,90000,101000,91000,73000,93000,109000,114000,140000,140000,145000,76000,154000,112000,129000,155000,132000,147000,124000)
     2. month <- c(1:24)
     3. plot(month,TableFarm\_Combined)
     4. abline(v= 13, col = 'red')
     5. legend("topleft", legend=c("Start of New Marketing"),col = 'red', lwd=1)
     6. 
     7. TableFarm\_marketing <- c("OldMarketing","OldMarketing","OldMarketing","OldMarketing","OldMarketing","OldMarketing","OldMarketing","OldMarketing","OldMarketing","OldMarketing","OldMarketing","OldMarketing","NewMarketing","NewMarketing","NewMarketing","NewMarketing","NewMarketing","NewMarketing","NewMarketing","NewMarketing","NewMarketing","NewMarketing","NewMarketing","NewMarketing")
     8. boxplot(TableFarm\_Combined ~ TableFarm\_marketing)
     9. 
     10. As we can see the marketing campaign has done well with the exception of one extreme outlier that is low for even the original data set.
   1. It seems that it would be possible to conclude that the marketing campaign would be a success as even to one standard deviation below the new mean is still above two standard deviations from the old one.
      1. Null Hypothesis: The new Marketing campaign does change the mean revenue monthly.
      2. Alt. Hypothesis: The new Marketing campaign does not change the mean revenue monthly.
   2. Since we are testing to see if there is any affect from the data, positive or negative we will run a 2 sample T test. This will let us know if there is any significant difference between the 2 means.
      1. TableFarm\_current <- c(103000,108000,101000,108000,89000,101000,90000,101000,91000,73000,93000,109000)
      2. TableFarm\_future <- c(114000,140000,140000,145000,76000,154000,112000,129000,155000,132000,147000,124000)
      3. t.test(TableFarm\_current,TableFarm\_future, paired = T, alternative = "two.sided")
         1. Paired t-test
         2. data: TableFarm\_future and TableFarm\_current
         3. t = 5.131, df = 11, p-value = 0.0003278
         4. alternative hypothesis: true difference in means is not equal to 0
         5. 95 percent confidence interval:
         6. 19082.18 47751.15
         7. sample estimates:
         8. mean of the differences
         9. 33416.67
      4. To the .001 significance level there is enough evidence to claim that the mean amount of income increased from the original data.
2. Can we detect an association between chocolate consumption and Nobel prizes?
   * 1. nobel <- read.csv("~/GitHub/ds710fall2016assignment6/nobel.csv",header = TRUE)
     2. attach(nobel)
     3. nobelwinners <- ne
     4. chocolateconsumtion <- c
     5. plot(chocolateconsumtion,nobelwinners)
     6. 
     7. m1 = lm(nobelwinners~chocolateconsumtion)
     8. summary(m1)
     9. 
     10. Linear model == n = -0.57710 + (c\*0.38545)
     11. This model compared to the original of n= -.8+(c\*0.4), is fairly similar the y intercept is a little higher than in the original, and the coefficient for chocolate is a little less.
     12. 
     13. Null Hypothesis: There is a correlation between chocolate consumption per-capita and the number of Nobel Awards received.
     14. Alternative Hypothesis: There is no correlation between the amount of chocolate consumed by a population per-capita and the amount of Nobel Awards received by that population.
   1. With a p value being less than the confidence interval of 0.001 we cannot reject the null Hypothesis that it is possible that there is correlation between the amount of chocolate consumed and number of Nobel awards received.
      1. plot(m1)
      2. 
      3. This plot shows a slight parabola telling us that there may be some non-linear relationship between the variables
      4. 
      5. The plot shows us that our residuals are fairly normally distributed because all of the points seem to follow a straight line.
      6. 
      7. The line in this chart show that it is likely that there is not equal variance in the data.
3. In homework 5, you counted the frequencies of letters in two encrypted texts.  In this problem, you will use statistical analysis to identify the language in which the text was written, and decrypt it.
   * 1. encryptedA <- read.csv("~/GitHub/ds710fall2016assignment6/encryptedA.csv",header = TRUE)
     2. attach(encryptedA)
     3. encrypt\_order <- order(value)
     4. barplot( value[encrypt\_order], names.arg = key[encrypt\_order] )
     5. 
     6. Freq <- read.csv("~/GitHub/ds710fall2016assignment6/Letter Frequencies.csv",header = TRUE)
     7. attach(Freq)
     8. par( mfrow= c(2,1))
     9. freq\_order <- order(English)
     10. barplot( value[encrypt\_order], names.arg = key[encrypt\_order], main = 'Encrypted A' )
     11. barplot( English[freq\_order], names.arg = Letter[freq\_order], main = 'English Frequency' )
     12. 
   1. It’s possible that the encrypted text came from English mainly because there is a distinct step down from the first letter and then at about the 8th or 9th letter. I would like to do more testing before deciding one way or another.
      1. ordered\_encrypta <- value[encrypt\_order]
      2. ordered\_freq <- English[freq\_order]
      3. ordered\_encrypta <- c( rep(0, 6), ordered\_encrypta )
      4. Null Hypothesis: The frequency of letters in Encrypted A does correlate to the frequency of letters in the English Language.
      5. Alternative Hypothesis: The frequency of letters in Encrypted A do not correlate to the frequency of letters in the English Language.
      6. encrypta\_mulitfreq <- ordered\_freq\*sum(ordered\_encrypta)
      7. sortEnglish\_combined = c( sum(encrypta\_mulitfreq[1:4]), encrypta\_mulitfreq[5:26] )
      8. sortEnglish\_combined\_prob = c( sum(ordered\_freq[1:4]), ordered\_freq[5:26] )
      9. sortencrypta\_combined = c( sum(ordered\_encrypta[1:4]), ordered\_encrypta[5:26] )
      10. chisq.test(sortencrypta\_combined, p=sortEnglish\_combined\_prob)
      11. 
   2. The test results are significant to the 0.01 significant level. Which means that the encrypted data is likely encrypted from the English language.
      1. #repeat for welsh in encryptA
         1. freq\_order\_W <- order(Welsh)
         2. ordered\_freq\_W<- Welsh[freq\_order\_W]
         3. welsh\_multifreq <- ordered\_freq\_W\*sum(ordered\_encrypta)
         4. sortWelsh\_combined = c( sum(encrypta\_mulitfreq[1:7]), encrypta\_mulitfreq[8:26] )
         5. sortWelsh\_combined\_prob = c( sum(ordered\_freq\_W[1:7]), ordered\_freq\_W[8:26] )
         6. sortencrypta\_combined\_W = c( sum(ordered\_encrypta[1:7]), ordered\_encrypta[8:26] )
         7. chisq.test(sortencrypta\_combined\_W, p=sortWelsh\_combined\_prob)
         8. 
      2. #repeat for english in encryptB
         1. encryptedB <- read.csv("~/GitHub/ds710fall2016assignment6/encryptedB.csv",header = TRUE)
         2. keyB <- encryptedB$key
         3. valueB <- encryptedB$value
         4. encryptB\_order <- order(valueB)
         5. ordered\_encryptB <- valueB[encryptB\_order]
         6. English\_multifreqB <- ordered\_freq\*sum(ordered\_encryptB)
         7. sortEnglish\_combinedB = c( sum(English\_multifreqB[1:4]), English\_multifreqB[5:26] )
         8. sortEnglish\_combinedB\_prob = c( sum(ordered\_freq[1:4]), ordered\_freq[5:26] )
         9. sortencryptB\_combined = c( sum(ordered\_encryptB[1:4]), ordered\_encryptB[5:26] )
         10. chisq.test(sortencryptB\_combined, p=sortEnglish\_combinedB\_prob)
         11. 
      3. #repeat for welsh in encryptB
         1. welsh\_multifreqB <- ordered\_freq\_W\*sum(ordered\_encryptB)
         2. sortWelsh\_combinedB = c( sum(welsh\_multifreqB[1:7]), welsh\_multifreqB[8:26] )
         3. sortWelsh\_combinedB\_prob = c( sum(ordered\_freq\_W[1:7]), ordered\_freq\_W[8:26] )
         4. sortencryptB\_combined\_Welsh = c( sum(ordered\_encryptB[1:7]), ordered\_encryptB[8:26] )
         5. chisq.test(sortencryptB\_combined\_Welsh, p=sortWelsh\_combinedB\_prob)
         6. 
      4. Based on the four Chi squared test I believe it is safe to say that encryptA is from English and encryptB is from Welsh. The encryptB looked as if it may also come from English as the p-value for that test was close to the .05 confidence level although still greater that it. But once the test for Welsh was done and the results returned it was obviously more related to Welsh with a p-value lest than 2.2e-16.