S&DS 230 Final Project - Fluffy Unicorns

Due by 11:59pm, Saturday, May 6, 2023

S&DS 230/530/ENV 757

Introduction

Our motivation with this project is to investigate college students' varying relationships with food. We used a number of questions that go along with this to identify these relationships, from their life satisfaction and gender to eating habits and GPA. Because many of the variables we used don't intuitively have high degrees of correlation, we did not expect to find much significance in our results, but some interesting results were found between variables.

Data

Within the data set, there are about 60 variables. However, we use 10 variables in our analysis: GPA (on a 4.0 scale), weight (in lbs), frequency of exercise (ordinal, on a scale from never to everyday, 1 indicating higher frequency), frequency of cooking (ordinal, on a scale from never to everyday, 1 indicating higher frequency), life satisfaction (ordinal, from 1-10, 1 indicates most satisfaction), gender (1 being female and 2 being male), frequency of checking nutritional levels (ordinal, 5 indicating higher frequency), change in eating habits in college (with recoding, ordinal, with a higher number indicating better habits), self-perceived weight (ordinal, with higher number indicating higher perceived weight), and healthy feeling (ordinal, on a scale of 1 to 10, 1 indicating strong agreement with the statement "I feel very healthy").

First, we downloaded packages necessary for our analysis:

```
library(car)
library(leaps)
library(lubridate)
library(rvest)
library(corrplot)
library(knitr)
source("http://www.reuningscherer.net/s&ds230/Rfuncs/regJDRS.txt")
```

Data Cleaning

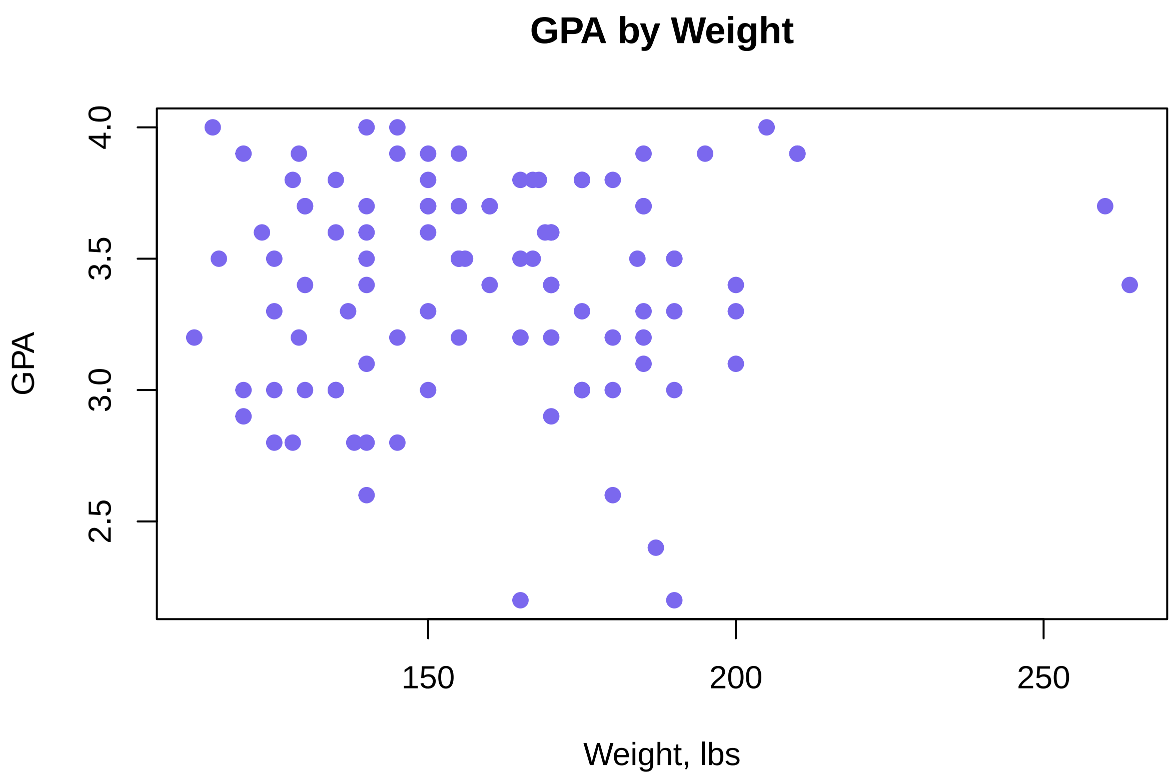
The next thing we did was clean the data. To keep data consistent, we rounded all GPAs to one decimal point and removed all commentary such that we could run numerical tests on GPA. We did the same for weight, though we did not round. We also recoded Gender into "F" and "M" rather than leaving them with the numeric indicators. Further, we recoded the exercise frequency and cooking frequency such that a higher number indicated higher frequency. Finally, we removed all rows with NAs.

```
food <- read.csv("Food.csv")
food <- food[c(1, 2, 11, 18, 23, 34, 41, 46, 51, 61)]

#cleaning GPA
gpa <- food$GPA
gpa <- gsub(".* .*", "",gpa)
gpa <- trimws(gpa)
```

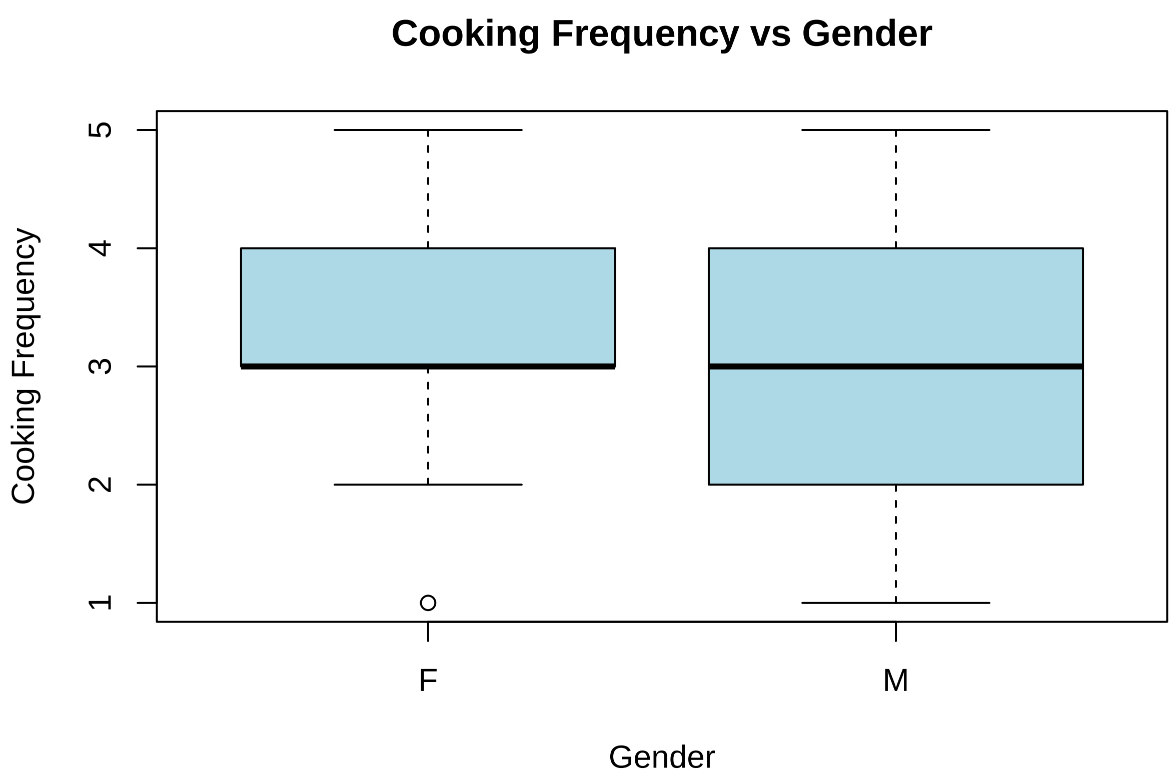
```r
gpa <- as.numeric(gpa)
gpa <- round(gpa, 1)
food$GPA <- gpa

#cleaning weight
food$weight <- gsub("[^0-9]", NA, food$weight)
food$weight <- as.numeric(food$weight)

#recoding Gender
food$Gender <- recode(food$Gender, "1 = 'F'; 2 = 'M'")

#recoding eating changes
food$eating_changes_coded <- recode(food$eating_changes_coded, "1 = '1'; 2 = '3'; 3 = '2'")
food$eating_changes_coded[food$eating_changes_coded == 4] <- NA
food$eating_changes_coded <- as.numeric(food$eating_changes_coded)

#convert ambiguity in self-perceived weight to NA
food$self_perception_weight[food$self_pereception_weight == 6] <- NA

#reworking frequency
food$exercise <- 6 - food$exercise
food$cook <- 6 - food$cook

#remove NA's
food <- na.omit(food)

attach(food)
```

Graphics

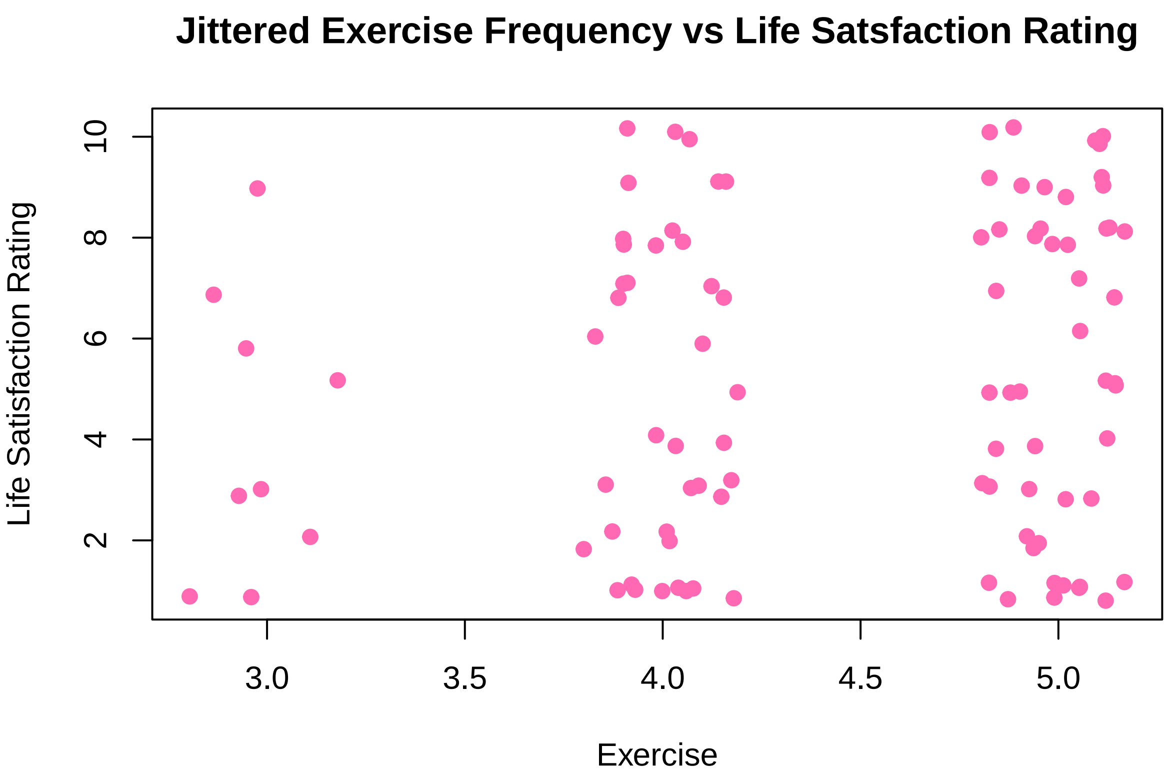
```r
plot(weight, GPA, pch = 19, main = "GPA by Weight", col = "mediumslateblue", xlab = "Weight, lbs", ylab
```




```
boxplot(food$cook ~ food$Gender, col = "lightblue", xlab = "Gender", ylab = "Cooking Frequency", main =
```



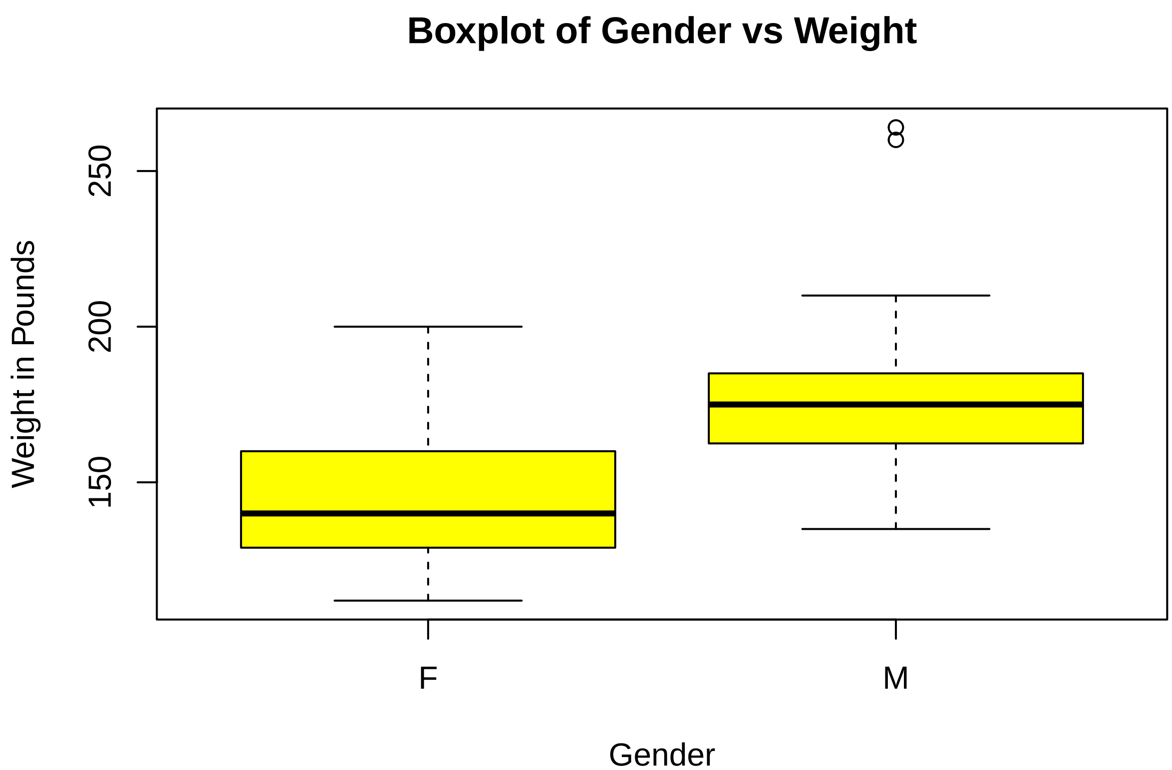

```
plot(jitter(exercise), jitter(life_rewarding), pch = 19, main = "Jittered Exercise Frequency vs Life Sa
```



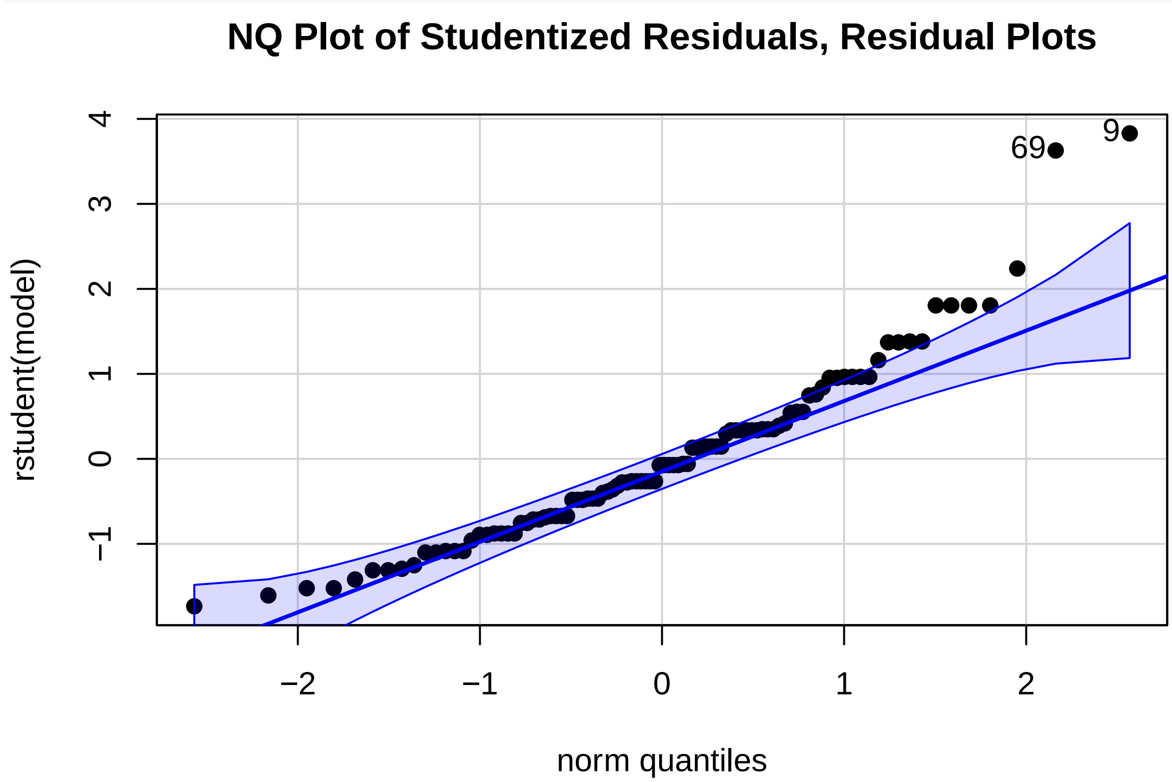
In this section we include multiple insightful graphics to display relationships between different variables visibly. As an additional note, there are more graphics included in other sections. Our first graph is a scatterplot of weight and GPA; there seems to be no obvious relationship between the two. Our second graph is a boxplot of how often people cook their own meals, separated by gender. Both groups have a median of 3 (the middle value), but males have a larger interquartile range and range. Our third graph is a jittered plot of exercise versus life satisfaction; again there does not seem to be a significant relationship between amount exercised and life satisfaction.

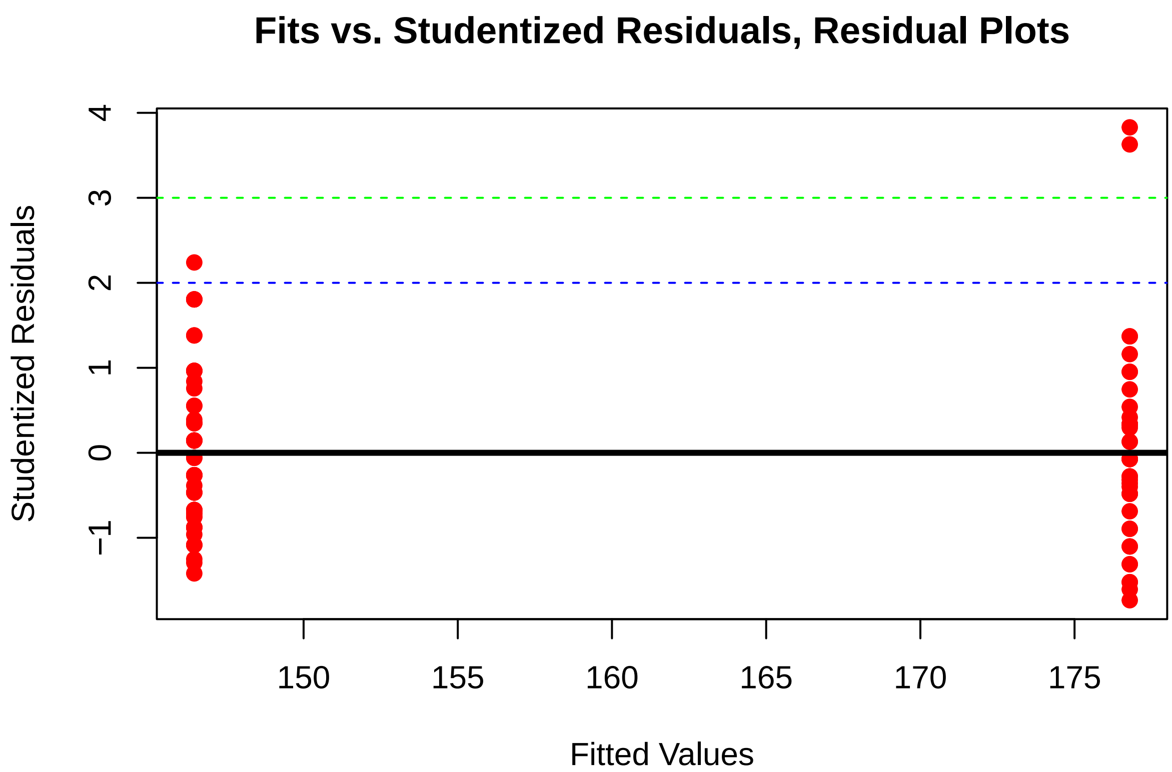
Basic Tests

```
#boxplot Gender vs GPA
boxplot(food$weight ~ food$Gender, main = "Boxplot of Gender vs Weight", col = "yellow", ylab = "Weight
```




```
#ensure assumptions for weight and gender are met - not perfectly ideal bc there are two outliers
lm1 <- lm(food$weight ~ food$Gender)
myResPlots2(lm1)
```

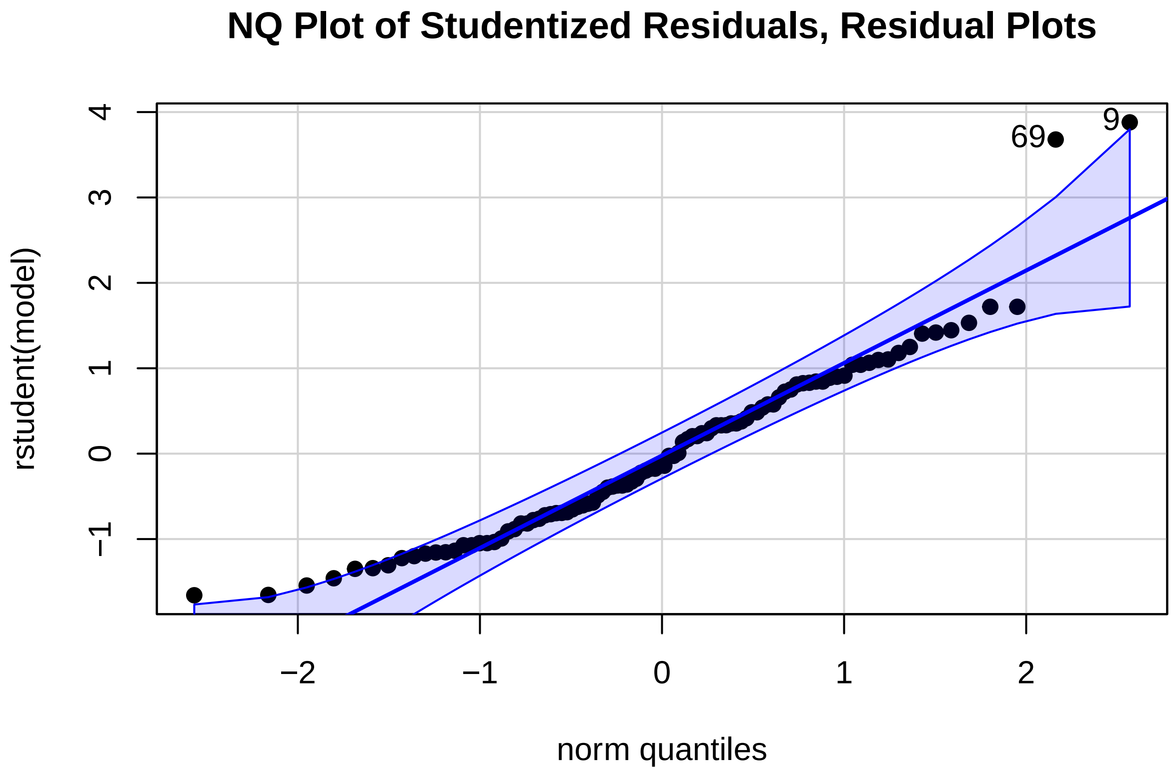


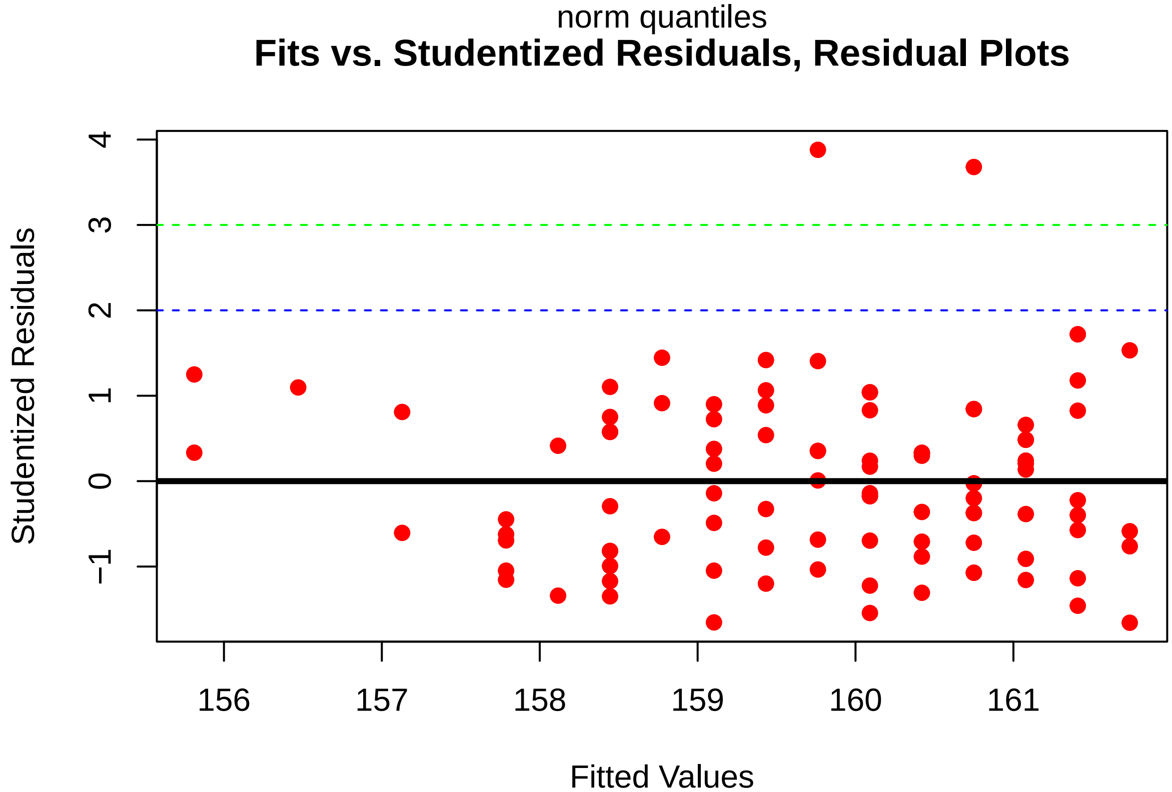



```r
#t.test for weight and gender
t.test(food$weight ~ food$Gender)
```

```
##
##  Welch Two Sample t-test
##
## data:  food$weight by food$Gender
## t = -5.9091, df = 80.713, p-value = 7.852e-08
## alternative hypothesis: true difference in means between group F and group M is not equal to 0
## 95 percent confidence interval:
##  -40.55142 -20.12089
## sample estimates:
## mean in group F mean in group M
##        146.4545        176.7907
```

```r
#ensure assumptions for weight and GPA are met - not perfectly ideal but pretty good bc only 2 outliers
lm2 <- lm(food$weight ~ food$GPA)
myResPlots2(lm2)
```






```r
#find 95% conf interval for the correlation for between gender and weight
cor.test(food$GPA, food$weight)
```

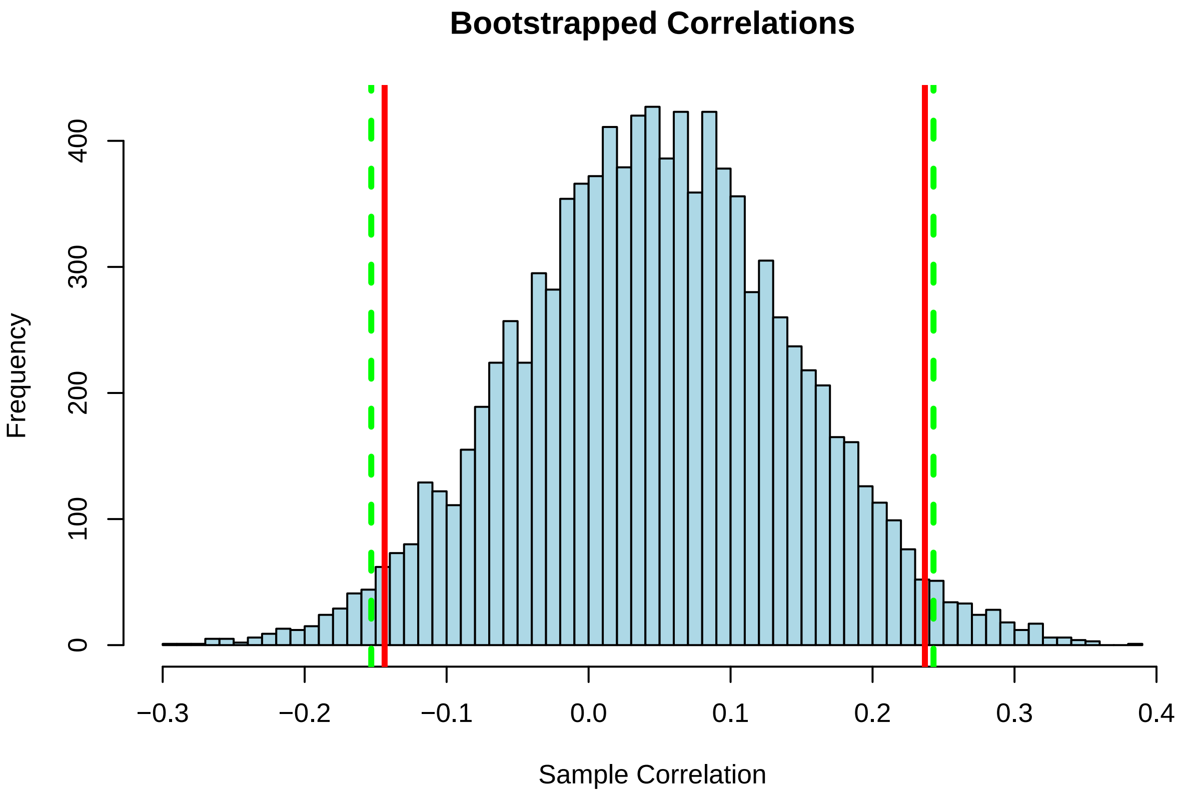
```
##
##  Pearson's product-moment correlation
##
```

```
## data:  food$GPA and food$weight
## t = 0.45861, df = 96, p-value = 0.6475
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.1530854  0.2429227
## sample estimates:
##        cor
## 0.04675562
```

```
#bootstrap confidence interval
N <- nrow(food)
n_samp <- 10000
corres <- rep(NA, n_samp)
for(i in 1:n_samp){
  s <- sample(1:N, N , replace = T)
  fakeData <- food[s, ]
  corres[i] <- cor(fakeData[, 1], fakeData[, 10])
}

bci <- quantile(corres, c(0.025, 0.975))
bci
```

```
##       2.5%      97.5%
## -0.1436711  0.2368686
```

```
#graph demonstrating bootstrap vs sample correlations!
par(cex = 0.8)
hist(corres, col = "lightblue", main = "Bootstrapped Correlations", xlab = "Sample Correlation", breaks
abline(v = bci, lwd = 3, col = "red")
abline(v = cor.test(food$GPA, food$weight)$conf.int, lwd = 3, col = "green", lty = 2)
legend(-0.36, 600, c("Theoretical CI","Boot CI"), lwd = 3, col = c("green","red"), lty = c(2, 1))
```



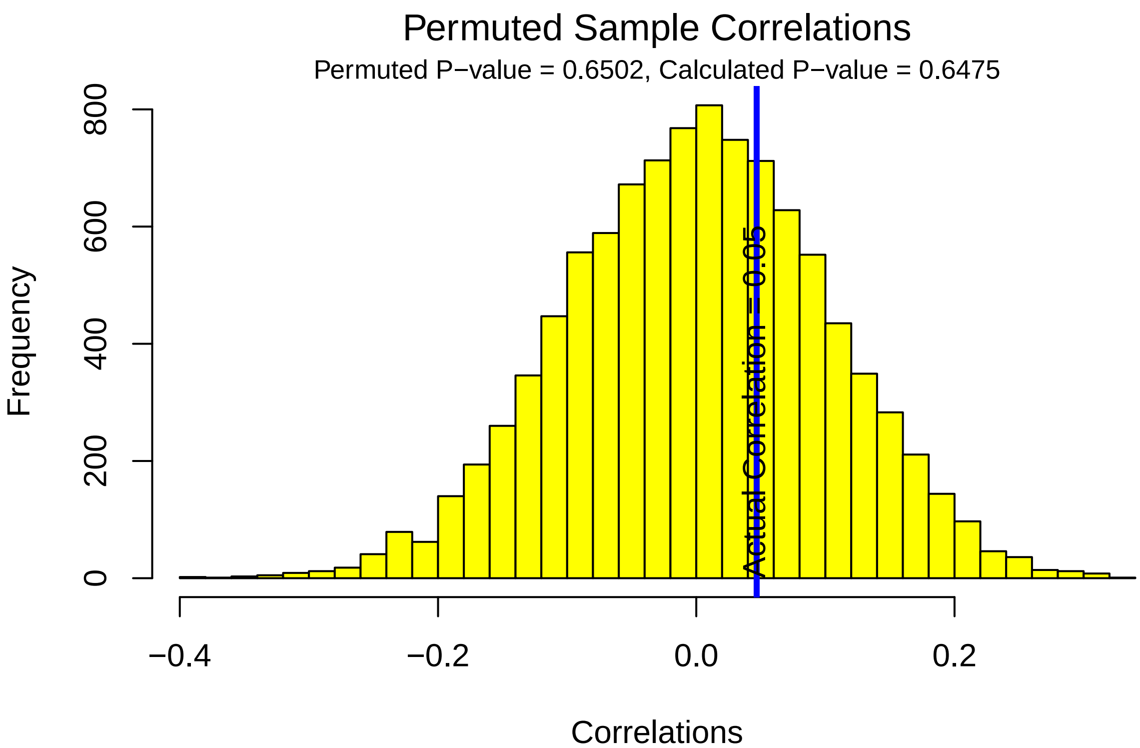
Next, we ran several basic tests on multiple variables. We started by looking at a boxplot of weight in pounds, separated by gender to see if there were observable differences. Then, before running a t-test to examine the significance of this relationship, we ensured that the data met the necessary assumptions. Since the data for variables Gender and Weight mostly fall within the bounds of normal distribution and the residual plots don't demonstrate any worrying heteroskedasticity BUT there are two large upper outliers, we proceed with caution. The t-test reveals that 0 is not in the confidence interval; there is a statistically significant difference between weights for different genders. Mean female weight is 146.5lbs and mean male weight is 176.8lbs. Next, we run tests on the relationship between weight and GPA; again, the data mostly fall within the bounds of normal distribution and the residual plots don't demonstrate any worrying heteroskedasticity BUT there are two large upper outliers. We proceed with caution. Our correlation test reveals a confidence interval that DOES include 0, so there is not a statistically signficant correlation. Our bootstrapped test for the same correlation using 10,000 samples reveals the same thing (although the interval is slightly narrower) that there is not a statistically significant correlation between GPA and weight.

Permutation Test

We also decided to run a permutation test on GPA vs Weight, because we were unsure about the normality of the data sets, and permutation tests are effective whenwe do not know the population distribution of the variables. The results are below:

```
#permutation test between GPA and weight
permCor <- function(x, y, n_samp = 10000, plotit = T){
  corResults <- rep(NA, n_samp)
  for (i in 1:n_samp){
    corResults[i] <- cor(x, sample(y), use = "complete.obs")
  }

  pval <- mean(abs(corResults) >= abs(cor(x, y, use = "complete.obs")))
```

```r
  if (plotit == T){
    hist(corResults, col = "yellow", main = "", xlab = "Correlations", breaks = 50, xlim = range(corResu
      mtext("Permuted Sample Correlations", cex = 1.2, line = 1)
      mtext(paste0("Permuted P-value = ", round(pval, 4),", Calculated P-value = ", round(cor.test(x, y
      abline(v = cor(x, y), col = "blue", lwd = 3)
      text(cor(x, y, use = "complete.obs") * 0.95, 0, paste("Actual Correlation =", round(cor(x,y, use =
   }
  if (plotit == F){
    return(round(pval, 5))
  }
}

permCor(food$GPA, food$weight)
```



From the graph, we see an r-squared value of 0.05, and a calculated p-value of 0.6475. The large p-value shows that these variables are not statistically significantly correlated.

ANOVA

We chose to run an ANOVA test on the difference of life rewarding rating between the different exercise frequency groups. Since there were no responses with an exercise frequency of 1 or 2, we only examined the groups of frequency 3, 4, and 5. First, we calculate the standard deviations between the groups:

```r
(sds <- tapply(food$life_rewarding, food$exercise, sd))
```

```
##        3        4        5
## 2.803767 3.148806 3.183183
```

```r
round(max(sds)/min(sds), 1)
```

```
## [1] 1.1
```

Since the largest SD divided by the smallest SD is 1.1, which is less than 2, we can proceed with the ANOVA test:

```
(aov1 <- aov(food$life_rewarding ~ food$exercise))
```

```
## Call:
##    aov(formula = food$life_rewarding ~ food$exercise)
##
## Terms:
##                 food$exercise Residuals
## Sum of Squares        17.6055  936.3945
## Deg. of Freedom             1        96
##
## Residual standard error: 3.123157
## Estimated effects may be unbalanced
```

```
summary(aov1)
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## food$exercise  1   17.6  17.605   1.805  0.182
## Residuals     96  936.4   9.754
```

Since the p value of 0.182 is larger than alpha = 0.05 we cannot reject the null hypothesis, meaning there is no significant difference in life rewarding rating between the different exercise groups.

ANCOVA

Next, we attempt to identify the significance between the variables Gender and Education in predicting Weight through an ANCOVA test:

```
dim(food)
```

```
## [1] 98 10
```

```
Anvmod1 <- lm(weight ~ Gender*exercise)
#Again, get overall test of significance of terms
Anova(Anvmod1, type = 3)
```

```
## Anova Table (Type III tests)
##
## Response: weight
##                 Sum Sq Df F value   Pr(>F)
## (Intercept)      30450  1 50.6754 2.15e-10 ***
## Gender            1561  1  2.5986   0.1103
## exercise           268  1  0.4465   0.5056
## Gender:exercise    305  1  0.5082   0.4777
## Residuals        56484 94
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Specific Differences
summary(Anvmod1)
```

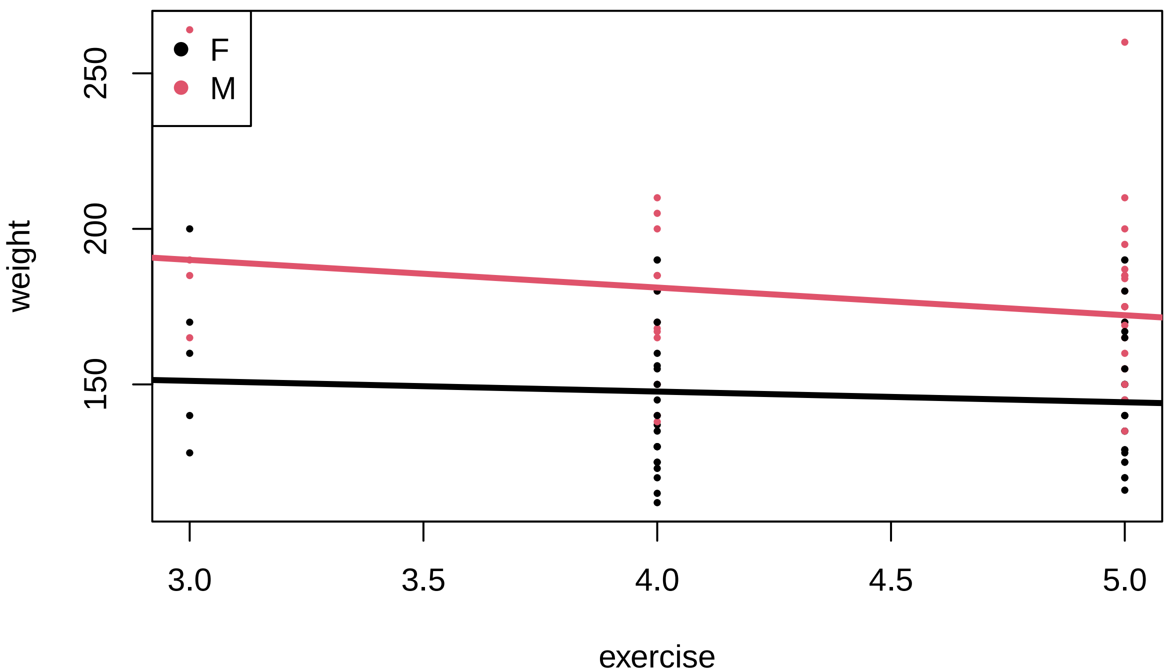
```
##
## Call:
## lm(formula = weight ~ Gender * exercise)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -43.132 -17.589  -2.474  12.161  87.757
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       161.448     22.680   7.119 2.15e-10 ***
## GenderM            55.237     34.266   1.612    0.110
## exercise           -3.436      5.142  -0.668    0.506
## GenderM:exercise   -5.452      7.648  -0.713    0.478
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24.51 on 94 degrees of freedom
## Multiple R-squared:  0.2978, Adjusted R-squared:  0.2754
## F-statistic: 13.29 on 3 and 94 DF,  p-value: 2.642e-07
```

```r
plot(weight ~ exercise, col=factor(Gender), pch=16, cex=.5)
legend("topleft", col = 1:2, legend = levels(factor(Gender)), pch = 16)
Ancoefs <- coef(Anvmod1)
Ancoefs
```

```
##      (Intercept)          GenderM         exercise GenderM:exercise
##       161.448000        55.236864        -3.436000        -5.452337
```

```r
abline(a = Ancoefs[1], b = Ancoefs[3], col = 1, lwd = 3)
abline(a = Ancoefs[1] + Ancoefs[2], b = Ancoefs[3] + Ancoefs[4], col = 2, lwd = 3)
```

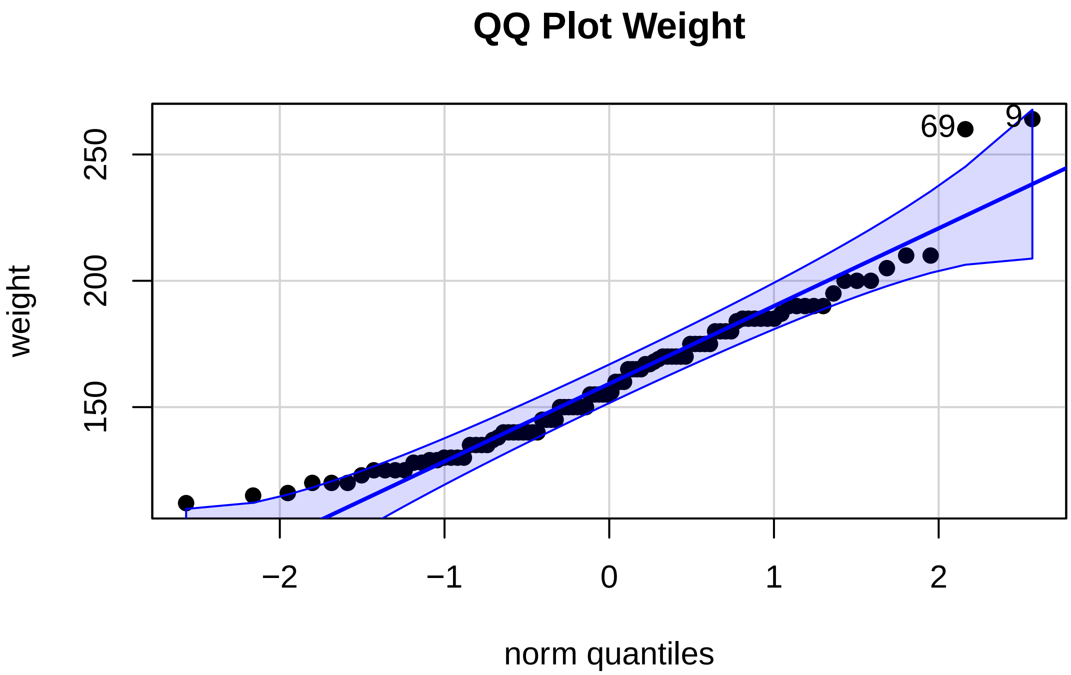


The p-value of 0.478, which is clearly greater than 0.05 or any other alpha-value we could reasonably expect, found by the ANCOVA test indicates that the interaction between Gender and Education is non-significant in predicting Weight. Below, we create a plot of Weight as predicted by Education with separate colors to indicate Gender, superimposing our two predicted regression lines for Males and Females. The lines clearly don't interact, proving consistent with our results from our initial test.

Multiple Regression

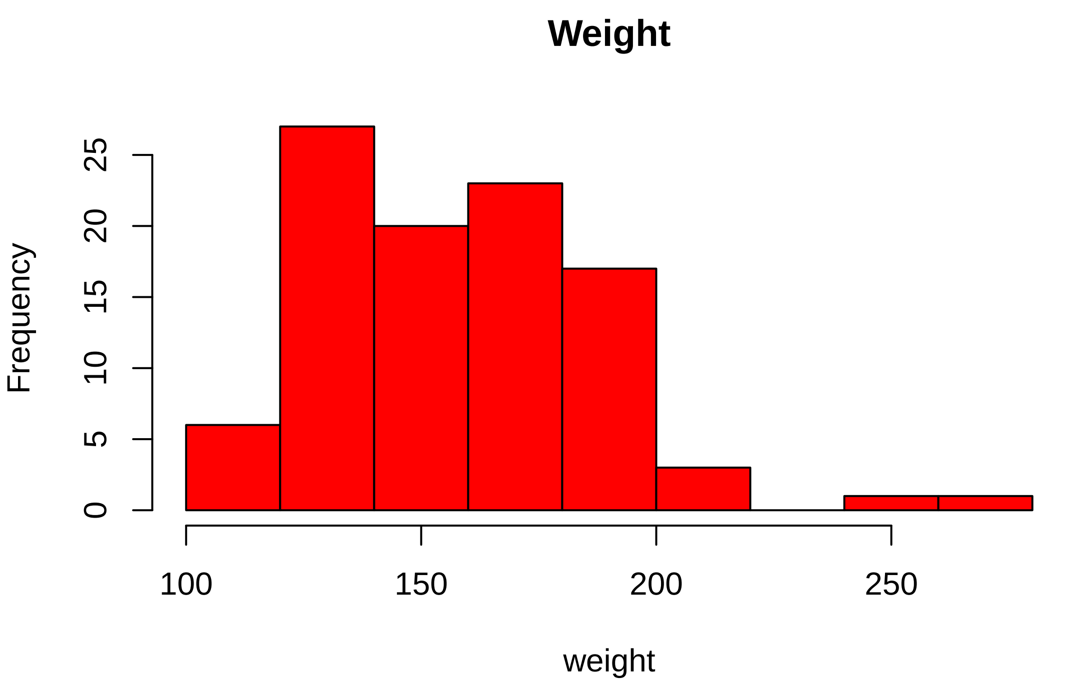
Finally, we ran a multiple regression to analyze the correlation between weight and life_rewarding, eating_changes_coded, and exercise:

```
qqPlot(weight, main = "QQ Plot Weight", pch = 19)
```



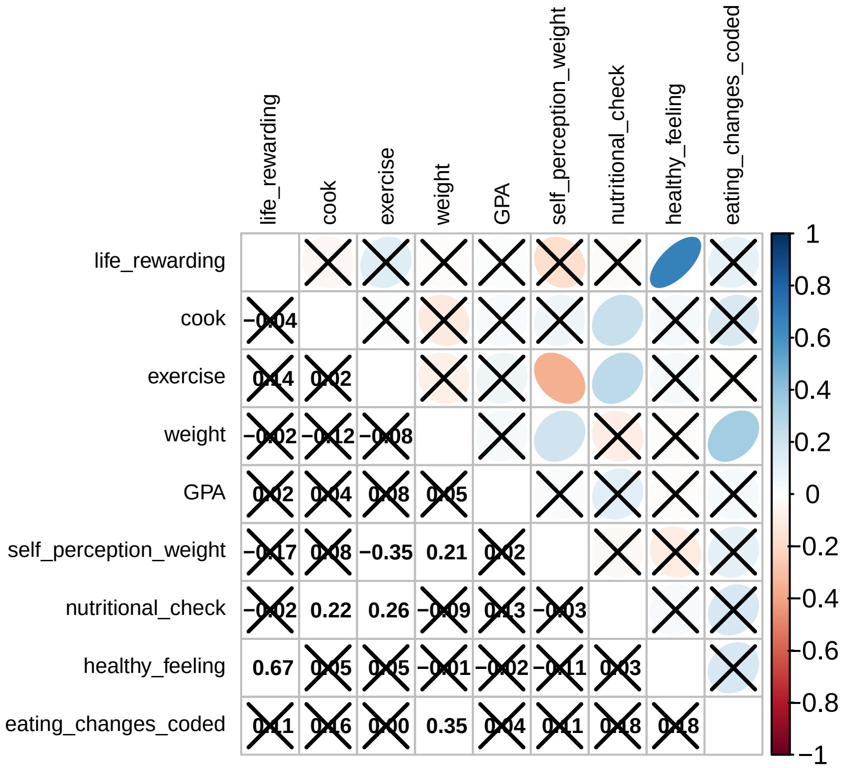

```
## [1]  9 69
```

```
hist(weight, col = "red", main = "Weight")
```

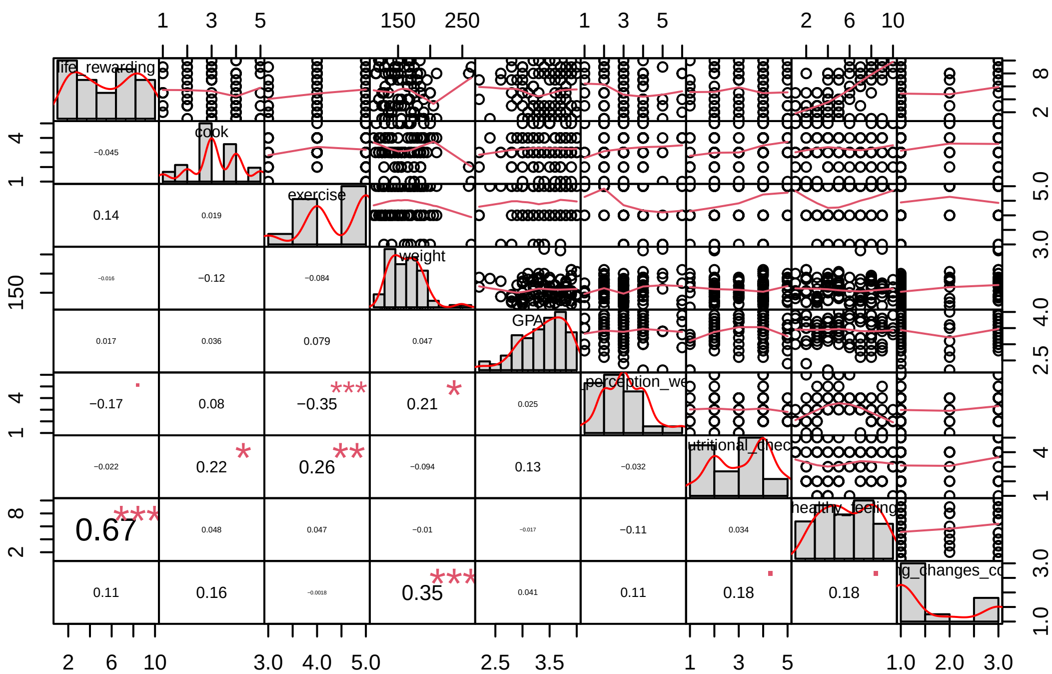



```
food <- na.omit(food[, c("life_rewarding", "cook", "exercise", "weight", "GPA", "self_perception_weight
```

```
sigcorr <- cor.mtest(food, conf.level = .95)
corrplot.mixed(cor(food), lower.col="black", upper = "ellipse", tl.col = "black", number.cex=.7,
               tl.pos = "lt", tl.cex=.7, p.mat = sigcorr$p, sig.level = .05)
```




```
pairsJDRS(food)
```



```
lm1 <- lm(weight ~ GPA + cook + exercise + self_perception_weight + nutritional_check + eating_changes_c
summary(lm1)
```

```
##
## Call:
## lm(formula = weight ~ GPA + cook + exercise + self_perception_weight +
##     nutritional_check + eating_changes_coded + healthy_feeling,
##     data = food)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -46.609 -18.232   0.486  16.641  68.919
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            139.5332    32.5665   4.285 4.58e-05 ***
## GPA                      3.3437     6.7052   0.499 0.619224
## cook                    -4.5647     2.7367  -1.668 0.098805 .
## exercise                 0.5837     4.5957   0.127 0.899216
## self_perception_weight   4.4225     2.6474   1.671 0.098294 .
## nutritional_check       -3.1829     2.5195  -1.263 0.209745
## eating_changes_coded    12.5331     3.2297   3.881 0.000198 ***
## healthy_feeling         -0.5052     1.0499  -0.481 0.631511
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.71 on 90 degrees of freedom
## Multiple R-squared:  0.2015, Adjusted R-squared:  0.1394
## F-statistic: 3.245 on 7 and 90 DF,  p-value: 0.004073
```

```
mod2 <- regsubsets(weight ~., data = food)
mod2sum <- summary(mod2)
mod2sum$which
```

```
##   (Intercept) life_rewarding  cook exercise   GPA self_perception_weight
## 1        TRUE          FALSE FALSE    FALSE FALSE                  FALSE
## 2        TRUE          FALSE  TRUE    FALSE FALSE                  FALSE
## 3        TRUE          FALSE  TRUE    FALSE FALSE                   TRUE
## 4        TRUE          FALSE  TRUE    FALSE FALSE                   TRUE
## 5        TRUE          FALSE  TRUE    FALSE  TRUE                   TRUE
## 6        TRUE          FALSE  TRUE    FALSE  TRUE                   TRUE
## 7        TRUE          FALSE  TRUE     TRUE  TRUE                   TRUE
## 8        TRUE           TRUE  TRUE     TRUE  TRUE                   TRUE
##   nutritional_check healthy_feeling eating_changes_coded
## 1             FALSE           FALSE                 TRUE
## 2             FALSE           FALSE                 TRUE
## 3             FALSE           FALSE                 TRUE
## 4              TRUE           FALSE                 TRUE
## 5              TRUE           FALSE                 TRUE
## 6              TRUE            TRUE                 TRUE
## 7              TRUE            TRUE                 TRUE
## 8              TRUE            TRUE                 TRUE
```

```
modnum <- which.max(mod2sum$rsq)
names(food)[mod2sum$which[modnum,]][-1]
```

```
## [1] "cook"                   "exercise"               "weight"
## [4] "GPA"                    "self_perception_weight" "nutritional_check"
## [7] "healthy_feeling"        "eating_changes_coded"
```

```r
foodtemp <- food[,mod2sum$which[modnum,]]
summary(lm(weight ~ ., data = foodtemp))
```

```
##
## Call:
## lm(formula = weight ~ ., data = foodtemp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -46.070 -18.387   0.465  17.453  69.080
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              139.4761    32.7478   4.259 5.08e-05 ***
## life_rewarding            -0.1689     1.2005  -0.141 0.888437
## cook                      -4.5971     2.7614  -1.665 0.099474 .
## exercise                   0.6614     4.6538   0.142 0.887314
## GPA                        3.3849     6.7484   0.502 0.617194
## self_perception_weight     4.3891     2.6725   1.642 0.104045
## nutritional_check         -3.2132     2.5425  -1.264 0.209602
## healthy_feeling           -0.3732     1.4123  -0.264 0.792180
## eating_changes_coded      12.5470     3.2490   3.862 0.000213 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.86 on 89 degrees of freedom
## Multiple R-squared:  0.2017, Adjusted R-squared:   0.13
## F-statistic: 2.811 on 8 and 89 DF,  p-value: 0.007968
```

```r
#Best model according to Adjusted R-squared
modnum <- which.max(mod2sum$adjr2)
names(food)[mod2sum$which[modnum, ]][-1]
```

```
## [1] "exercise"               "self_perception_weight" "nutritional_check"
## [4] "eating_changes_coded"
```

```r
foodtemp <- food[,mod2sum$which[modnum,]]
summary(lm(weight ~ ., data = foodtemp))
```

```
##
## Call:
## lm(formula = weight ~ ., data = foodtemp)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -46.13 -19.05  -0.14  16.17  77.64
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              137.6935    24.4483   5.632 1.92e-07 ***
## life_rewarding            -0.3075     0.8973  -0.343  0.73256
## exercise                   0.9469     4.6368   0.204  0.83865
## self_perception_weight     4.2123     2.6659   1.580  0.11753
## nutritional_check         -3.9408     2.4802  -1.589  0.11551
## eating_changes_coded      11.8033     3.2057   3.682  0.00039 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.89 on 92 degrees of freedom
## Multiple R-squared:  0.1732, Adjusted R-squared:  0.1282
## F-statistic: 3.853 on 5 and 92 DF,  p-value: 0.003247
```

```
#Best model according to Bayesian Information Criteria (BIC)
modnum <- which.min(mod2sum$bic)
names(food)[mod2sum$which[modnum, ]][-1]
```

```
## [1] "eating_changes_coded"
```

```
foodtemp <- food[,mod2sum$which[modnum,]]
summary(lm(weight ~ ., data = foodtemp))
```

```
##
## Call:
## lm(formula = weight ~ ., data = foodtemp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -48.888 -20.882  -1.001  17.081  90.136
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          143.7184     7.0162  20.484  < 2e-16 ***
## life_rewarding        -0.5030     0.8873  -0.567 0.572137
## eating_changes_coded  11.5576     3.1589   3.659 0.000416 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.24 on 95 degrees of freedom
## Multiple R-squared:  0.1237, Adjusted R-squared:  0.1053
## F-statistic: 6.707 on 2 and 95 DF,  p-value: 0.001885
```

```
#Best model according to Cp Statistic (a bit more complicated)
modnum <- min(c(1:length(mod2sum$cp))[mod2sum$cp < c(1:length(mod2sum$cp))+1])
names(food)[mod2sum$which[modnum, ]][-1]
```

```
## [1] "exercise"             "eating_changes_coded"
```

```
foodtemp <- food[,mod2sum$which[modnum,]]
summary(lm(weight ~ ., data = foodtemp))
```

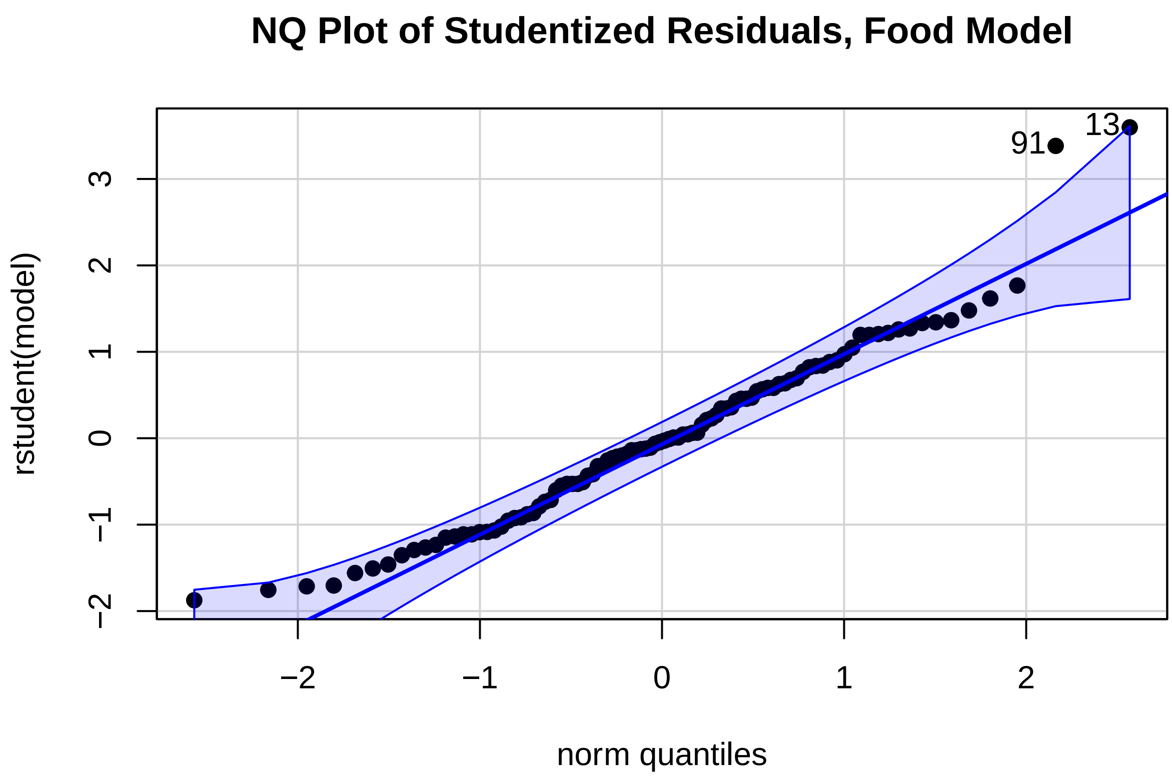
```
##
## Call:
## lm(formula = weight ~ ., data = foodtemp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -46.809 -20.448  -0.635  16.969  87.376
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          158.1943    19.5792   8.080 2.17e-12 ***
## life_rewarding        -0.4057     0.8975  -0.452 0.652304
```

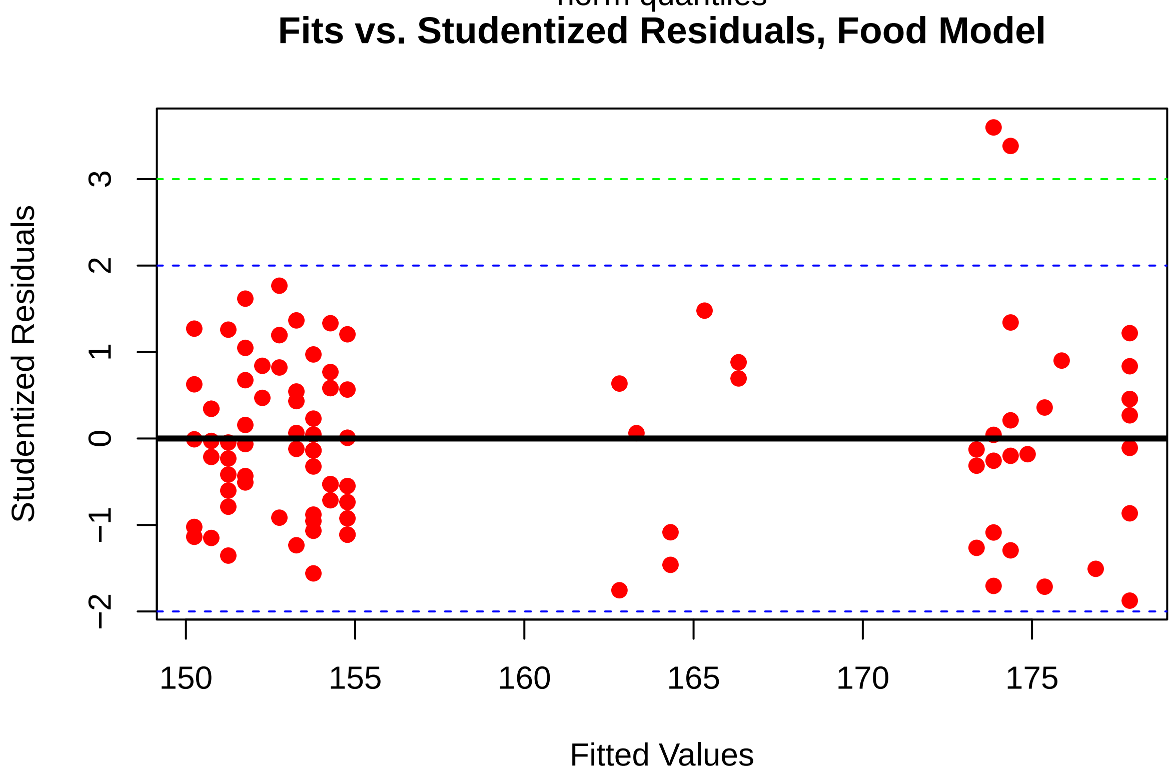
```
## exercise              -3.3740     4.2592  -0.792 0.430256
## eating_changes_coded  11.5149     3.1656   3.638 0.000449 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.29 on 94 degrees of freedom
## Multiple R-squared:  0.1295, Adjusted R-squared:  0.1018
## F-statistic: 4.663 on 3 and 94 DF,  p-value: 0.004406
```

```
#Final Model (BIC)
modnum <- which.min(mod2sum$bic)
foodtemp <- food[, mod2sum$which[modnum,]]
modfin <- lm(weight ~ ., data = foodtemp)
summary(modfin)
```

```
##
## Call:
## lm(formula = weight ~ ., data = foodtemp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -48.888 -20.882  -1.001  17.081  90.136
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          143.7184     7.0162  20.484  < 2e-16 ***
## life_rewarding        -0.5030     0.8873  -0.567 0.572137
## eating_changes_coded  11.5576     3.1589   3.659 0.000416 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.24 on 95 degrees of freedom
## Multiple R-squared:  0.1237, Adjusted R-squared:  0.1053
## F-statistic: 6.707 on 2 and 95 DF,  p-value: 0.001885
```

```
myResPlots2(modfin, "Food Model")
```





Finally, we did a multiple regression to analyze the correlation between weight and the other variables in our data set. The coefficients show that eating_changes_coded is a statistically significant predictor of weight, as the p value is 0.000416, while other all other predictors had p-values above .05, and so in the process of determining the best model were removed. The coefficient for eating_changes_coded is positive, indicating that as the level of eating changes increases, so does weight. This makes sense as a change in eating habits can often lead to weight fluctuation. The R-squared of the regression is .1237, indicating that the model explains

12.37% of the variance in the response variable, which is weight. This shows that this is a somewhat weak model, which given that most variables were insignificant, makes sense.

Conclusion

Throughout this project, we analyzed college students' relationships with food through a variety of survey questions. We looked at a few basic graphs, which showed us that many of the variables had little correlation. Next, we ran a t-test to examine the difference between weight between gender groups, and found a significant difference, with a mean female weight of 146.5lbs and mean male weight of 176.8lbs. Next, we calculated a confidence interval and bootstrapped confidence interval on the relationship between weight and GPA. Both intervals contained 0, meaning there was not a statistically significant correlation. We also ran a permutation test on GPA vs Weight, because we were unsure about the normality of the data sets. We got an r-squared value of 0.05, and a calculated p-value of 0.6475 showing again that the variables were not statistically significantly correlated. We then ran an ANOVA test on the difference of life rewarding rating between the different exercise frequency groups, and an ANCOVA test on the effect of Gender and Education in predicting Weight. We got p-values of 0.182 and 0.478 respectively, showing neither relationship was statistically significant. Finally, we ran a multiple regression to analyze the correlation between weight and the other variables. There was a significant correlation between eating habits changing in college and weight, but overall, the model accounted for only 12.37% of the variance in weight.