Module 3 Team Assignment

Team Assignment Expectations

This is a team assignment. Team assignments should be completed by each team as a group effort. There are expectations that all students will contribute their complete participation in the team assignments. Members of different teams are not to discuss these assignments with each other. If your team has questions or needs clarification, please contact me.

Details

Coffee and Coding Analysis of Variance and Design of Experiment

A dataset, provide by Kaggle, includes 9 variables – listed below. The dataset, including 100 observations is given in CofeeAndCode.csv.

```
library(HH)
```

```
## Loading required package: lattice

## Loading required package: grid

## Loading required package: latticeExtra

## Loading required package: multcomp

## Loading required package: mvtnorm

## Loading required package: survival

## Loading required package: TH.data

## Loading required package: MASS

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##     geyser

## Loading required package: gridExtra
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:gridExtra':
##
##     combine

## The following object is masked from 'package:MASS':
##
##     select

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
#library(reticulate)
#you may also need other packages

#set working directory

cc.data <- read.csv("CoffeeAndCode.csv", header = TRUE)
head(cc.data)
```

```
##   ï..CodingHours CoffeeCupsPerDay    CoffeeTime CodingWithoutCoffee
## 1              8                2 Before coding                 Yes
## 2              3                2 Before coding                 Yes
## 3              5                3  While coding                  No
## 4              8                2 Before coding                  No
## 5             10                3  While coding           Sometimes
## 6              8                2  While coding           Sometimes
##     CoffeeType CoffeeSolveBugs Gender Country AgeRange
## 1 CaffÃ¨ latte       Sometimes Female Lebanon 18 to 29
## 2    Americano             Yes Female Lebanon 30 to 39
## 3      Nescafe             Yes Female Lebanon 18 to 29
## 4      Nescafe             Yes   Male Lebanon     <NA>
## 5      Turkish              No   Male Lebanon 18 to 29
## 6      Nescafe             Yes   Male Lebanon 30 to 39
```

```
str(cc.data)
```

```
## 'data.frame':    100 obs. of  9 variables:
##  $ ï..CodingHours     : int  8 3 5 8 10 8 5 10 10 10 ...
##  $ CoffeeCupsPerDay   : int  2 2 3 2 3 2 2 4 2 2 ...
##  $ CoffeeTime         : chr  "Before coding" "Before coding" "While coding" "Before coding" ...
##  $ CodingWithoutCoffee: chr  "Yes" "Yes" "No" "No" ...
```

```
##  $ CoffeeType          : chr  "CaffÃ¨ latte" "Americano" "Nescafe" "Nescafe" ...
##  $ CoffeeSolveBugs     : chr  "Sometimes" "Yes" "Yes" "Yes" ...
##  $ Gender              : chr  "Female" "Female" "Female" "Male" ...
##  $ Country             : chr  "Lebanon" "Lebanon" "Lebanon" "Lebanon" ...
##  $ AgeRange            : chr  "18 to 29" "30 to 39" "18 to 29" NA ...
```

```
summary(cc.data)
```

```
##  ï..CodingHours  CoffeeCupsPerDay  CoffeeTime        CodingWithoutCoffee
##  Min.   : 1.00   Min.   :1.00     Length:100         Length:100
##  1st Qu.: 4.00   1st Qu.:2.00     Class :character   Class :character
##  Median : 7.00   Median :2.50     Mode  :character   Mode  :character
##  Mean   : 6.41   Mean   :2.89
##  3rd Qu.: 8.00   3rd Qu.:4.00
##  Max.   :10.00   Max.   :8.00
##   CoffeeType        CoffeeSolveBugs       Gender            Country
##  Length:100         Length:100         Length:100         Length:100
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##    AgeRange
##  Length:100
##  Class :character
##  Mode  :character
##
##
##
```

```
anyNA(cc.data)
```

```
## [1] TRUE
```

```
cc.data$CoffeeTimeF <- as.factor(cc.data$CoffeeTime)
cc.data$CodingWithoutCoffeeF <-
  as.factor(cc.data$CodingWithoutCoffee)
cc.data$CoffeeTypeF <- as.factor(cc.data$CoffeeType)
cc.data$CoffeeSolveBugsF <- as.factor(cc.data$CoffeeSolveBugs)
cc.data$GenderF <- as.factor(cc.data$Gender)
cc.data$CountryF <- as.factor(cc.data$Country)
cc.data$AgeRangeF <- as.factor(cc.data$AgeRange)

head(cc.data)
```

```
##   ï..CodingHours CoffeeCupsPerDay    CoffeeTime CodingWithoutCoffee
## 1              8                2 Before coding                 Yes
## 2              3                2 Before coding                 Yes
## 3              5                3  While coding                  No
## 4              8                2 Before coding                  No
## 5             10                3  While coding           Sometimes
## 6              8                2  While coding           Sometimes
```

```
##     CoffeeType CoffeeSolveBugs Gender Country AgeRange   CoffeeTimeF
## 1 CaffÃ¨ latte       Sometimes Female Lebanon 18 to 29 Before coding
## 2    Americano             Yes Female Lebanon 30 to 39 Before coding
## 3      Nescafe             Yes Female Lebanon 18 to 29  While coding
## 4      Nescafe             Yes   Male Lebanon     <NA> Before coding
## 5      Turkish              No   Male Lebanon 18 to 29  While coding
## 6      Nescafe             Yes   Male Lebanon 30 to 39  While coding
##   CodingWithoutCoffeeF  CoffeeTypeF CoffeeSolveBugsF GenderF CountryF AgeRangeF
## 1                  Yes CaffÃ¨ latte        Sometimes  Female  Lebanon  18 to 29
## 2                  Yes    Americano              Yes  Female  Lebanon  30 to 39
## 3                   No      Nescafe              Yes  Female  Lebanon  18 to 29
## 4                   No      Nescafe              Yes    Male  Lebanon      <NA>
## 5            Sometimes      Turkish               No    Male  Lebanon  18 to 29
## 6            Sometimes      Nescafe              Yes    Male  Lebanon  30 to 39
```

```
str(cc.data)
```

```
## 'data.frame':    100 obs. of  16 variables:
##  $ ï..CodingHours      : int  8 3 5 8 10 8 5 10 10 10 ...
##  $ CoffeeCupsPerDay    : int  2 2 3 2 3 2 2 4 2 2 ...
##  $ CoffeeTime          : chr  "Before coding" "Before coding" "While coding" "Before coding" ...
##  $ CodingWithoutCoffee : chr  "Yes" "Yes" "No" "No" ...
##  $ CoffeeType          : chr  "CaffÃ¨ latte" "Americano" "Nescafe" "Nescafe" ...
##  $ CoffeeSolveBugs     : chr  "Sometimes" "Yes" "Yes" "Yes" ...
##  $ Gender              : chr  "Female" "Female" "Female" "Male" ...
##  $ Country             : chr  "Lebanon" "Lebanon" "Lebanon" "Lebanon" ...
##  $ AgeRange            : chr  "18 to 29" "30 to 39" "18 to 29" NA ...
##  $ CoffeeTimeF         : Factor w/ 7 levels "After coding",..: 4 4 7 4 7 7 7 4 7 7 ...
##  $ CodingWithoutCoffeeF: Factor w/ 3 levels "No","Sometimes",..: 3 3 1 1 2 2 3 2 3 3 ...
##  $ CoffeeTypeF         : Factor w/ 8 levels "American Coffee",..: 3 2 7 7 8 7 7 8 1 7 ...
##  $ CoffeeSolveBugsF    : Factor w/ 3 levels "No","Sometimes",..: 2 3 3 3 1 3 2 2 2 1 ...
##  $ GenderF             : Factor w/ 2 levels "Female","Male": 1 1 1 2 2 2 2 2 2 2 ...
##  $ CountryF            : Factor w/ 1 level "Lebanon": 1 1 1 1 1 1 1 1 1 1 ...
##  $ AgeRangeF           : Factor w/ 5 levels "18 to 29","30 to 39",..: 1 2 1 NA 1 2 NA 1 2 2 ...
```

```
summary(cc.data)
```

```
##  ï..CodingHours  CoffeeCupsPerDay  CoffeeTime        CodingWithoutCoffee
##  Min.   : 1.00   Min.   :1.00     Length:100         Length:100
##  1st Qu.: 4.00   1st Qu.:2.00     Class :character   Class :character
##  Median : 7.00   Median :2.50     Mode  :character   Mode  :character
##  Mean   : 6.41   Mean   :2.89
##  3rd Qu.: 8.00   3rd Qu.:4.00
##  Max.   :10.00   Max.   :8.00
##
##   CoffeeType        CoffeeSolveBugs       Gender            Country
##  Length:100         Length:100         Length:100         Length:100
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
```

```
##    AgeRange                           CoffeeTimeF CodingWithoutCoffeeF
##  Length:100         After coding           : 2   No       :19
##  Class :character   All the time           : 4   Sometimes:51
##  Mode  :character   Before and while coding: 4   Yes      :30
##                     Before coding          :25
##                     In the morning         : 3
##                     No specific time       : 1
##                     While coding           :61
##                  CoffeeTypeF  CoffeeSolveBugsF   GenderF      CountryF
##  Nescafe               :32   No       :27      Female:26   Lebanon:100
##  American Coffee       :23   Sometimes:43      Male  :74
##  Turkish               :19   Yes      :30
##  Espresso (Short Black): 8
##  Cappuccino            : 7
##  (Other)               :10
##  NA's                  : 1
##     AgeRangeF
##  18 to 29:60
##  30 to 39:29
##  40 to 49: 6
##  50 to 59: 1
##  Under 18: 2
##  NA's    : 2
##
```

Part 1

Do Lebanese programmers consume coffee above the normal average level (comparing to the average consumption in Lebanon which is 1.4 cups of coffee per day)?

The overall average of Coffee Cups Per Day is 2.9. Confirm this. Consider groups of Lebanese programmers, does any particular grouping of Lebanese programmers average 1.4 or less cups of coffee per day?

See Helpful Notes in the assignment document.

```
# factors only with no interactions
coffee.aov <- aov(CoffeeCupsPerDay ~
                  CoffeeTimeF + CodingWithoutCoffeeF + CoffeeTypeF
                  + CoffeeSolveBugsF + GenderF + AgeRangeF,
                  data=cc.data)
model.tables(coffee.aov, "means")
```

```
## Tables of means
## Grand mean
##
## 2.876289
##
##  CoffeeTimeF
##     After coding All the time Before and while coding Before coding
##              4.5            6                   4.25         2.042
## rep          2.0            4                   4.00        24.000
##     In the morning No specific time While coding
##                  1               4        2.932
## rep              3               1       59.000
```

```
##
##  CodingWithoutCoffeeF
##         No Sometimes    Yes
##      3.479     2.932  2.406
## rep 18.000    50.000 29.000
##
##  CoffeeTypeF
##     American Coffee Americano CaffÃ¨ latte Cappuccino Double Espresso (Doppio)
##                2.841     2.736        2.471      3.307                     5.306
## rep           23.000     2.000        5.000      7.000                     3.000
##     Espresso (Short Black) Nescafe Turkish
##                      2.996   2.372   3.245
## rep                  8.000  30.000  19.000
##
##  CoffeeSolveBugsF
##         No Sometimes    Yes
##      2.639     2.719  3.317
## rep 26.000    42.000 29.000
##
##  GenderF
##     Female  Male
##      2.865  2.88
## rep 26.000 71.00
##
##  AgeRangeF
##     18 to 29 30 to 39 40 to 49 50 to 59 Under 18
##        2.953    2.865     2.32    2.036    2.877
## rep   59.000   29.000     6.00    1.000    2.000
```

Part 1 Findings:

The table of means confirms the overall mean is approximately 2.9 cups of coffee per day for Lebanese programmers. The only group of Lebanese programmers that averages less than 1.4 cups of coffee per day is the group whose coffee time is in the morning. This group averages 1 cup of coffee per day and only has 3 observations.

Part 2

Consider the following two questions:

1. Are there significant differences between coding hours by gender in Lebanese programmers?
2. Are there significant differences between the number of cups of coffee drank per day by gender in Lebanese programmers?

For both questions 1 & 2, a one-way ANOVA by gender show a significant difference in averages by gender. Confirm this. However, some argue if other variables are considered the differences between gender are not significant and can be explained by differences in the other given variables. Do you agree or disagree with this statement? Explain why and support your decision with other ANOVA models, output and at least one meaningful graphic for each question.

See Helpful Notes in the assignment document.

1. Coding Hours

```
#coding hours by gender
coding1.aov <- aov(ï..CodingHours ~ GenderF,
                   data=cc.data)
anova(coding1.aov)
```

```
## Analysis of Variance Table
##
## Response: ï..CodingHours
##           Df Sum Sq Mean Sq F value  Pr(>F)
## GenderF    1  36.94  36.942   5.525 0.02075 *
## Residuals 98 655.25   6.686
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#no interaction term - all factors
coding2.aov <- aov(ï..CodingHours ~ CoffeeTimeF + CodingWithoutCoffeeF + CoffeeTypeF
                     + CoffeeSolveBugsF + GenderF + AgeRangeF,
                     data=cc.data)
anova(coding2.aov)
```

```
## Analysis of Variance Table
##
## Response: ï..CodingHours
##                       Df Sum Sq Mean Sq F value Pr(>F)
## CoffeeTimeF            6  18.25  3.0418  0.4571 0.8377
## CodingWithoutCoffeeF   2  22.19 11.0944  1.6671 0.1958
## CoffeeTypeF            7  52.09  7.4414  1.1182 0.3610
## CoffeeSolveBugsF       2  13.99  6.9951  1.0511 0.3547
## GenderF                1  28.19 28.1888  4.2359 0.0431 *
## AgeRangeF              4  47.48 11.8699  1.7837 0.1412
## Residuals             74 492.45  6.6547
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Two-way interaction term - all factors
coding3.aov <- aov(ï..CodingHours ~ (CoffeeTimeF + CodingWithoutCoffeeF + CoffeeTypeF
                     + CoffeeSolveBugsF + GenderF + AgeRangeF)^2,
                     data=cc.data)
anova(coding3.aov)
```

```
## Analysis of Variance Table
##
## Response: ï..CodingHours
##                                      Df  Sum Sq Mean Sq F value  Pr(>F)
## CoffeeTimeF                           6  18.251  3.0418  0.4999 0.80059
## CodingWithoutCoffeeF                  2  22.189 11.0944  1.8235 0.18861
## CoffeeTypeF                           7  52.090  7.4414  1.2231 0.33825
## CoffeeSolveBugsF                      2  13.990  6.9951  1.1497 0.33781
## GenderF                               1  28.189 28.1888  4.6331 0.04443 *
```

```
## AgeRangeF                                4  47.480 11.8699  1.9509 0.14328
## CoffeeTimeF:CodingWithoutCoffeeF         4  46.550 11.6376  1.9127 0.14972
## CoffeeTimeF:CoffeeTypeF                 11  83.829  7.6208  1.2525 0.32100
## CoffeeTimeF:CoffeeSolveBugsF             3  39.213 13.0711  2.1484 0.12776
## CoffeeTimeF:GenderF                      1   1.806  1.8057  0.2968 0.59225
## CoffeeTimeF:AgeRangeF                    4  18.107  4.5266  0.7440 0.57388
## CodingWithoutCoffeeF:CoffeeTypeF         6   6.519  1.0865  0.1786 0.97937
## CodingWithoutCoffeeF:CoffeeSolveBugsF    4  20.401  5.1003  0.8383 0.51781
## CodingWithoutCoffeeF:GenderF             2   7.418  3.7092  0.6096 0.55384
## CodingWithoutCoffeeF:AgeRangeF           2  32.198 16.0992  2.6460 0.09688 .
## CoffeeTypeF:CoffeeSolveBugsF             7  36.234  5.1763  0.8508 0.56061
## CoffeeTypeF:GenderF                      4  33.218  8.3046  1.3649 0.28320
## CoffeeTypeF:AgeRangeF                    2  24.787 12.3934  2.0370 0.15795
## CoffeeSolveBugsF:GenderF                 2   2.955  1.4777  0.2429 0.78677
## CoffeeSolveBugsF:AgeRangeF               2  21.940 10.9699  1.8030 0.19188
## GenderF:AgeRangeF                        1   1.675  1.6748  0.2753 0.60589
## Residuals                               19 115.601  6.0843
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Two-way interaction term - Remove CoffeeTimeF - highest p-value (0.80059)
coding4.aov <- aov(ï..CodingHours ~ (CodingWithoutCoffeeF + CoffeeTypeF
                   + CoffeeSolveBugsF + GenderF + AgeRangeF)^2,
                   data=cc.data)
anova(coding4.aov)
```

```
## Analysis of Variance Table
##
## Response: ï..CodingHours
##                                         Df  Sum Sq Mean Sq F value  Pr(>F)
## CodingWithoutCoffeeF                     2  22.963 11.4815  1.7793 0.18328
## CoffeeTypeF                              7  46.871  6.6959  1.0377 0.42265
## CoffeeSolveBugsF                         2  13.534  6.7668  1.0487 0.36086
## GenderF                                  1  22.118 22.1181  3.4277 0.07233 .
## AgeRangeF                                4  46.187 11.5467  1.7894 0.15236
## CodingWithoutCoffeeF:CoffeeTypeF        10  76.667  7.6667  1.1881 0.33072
## CodingWithoutCoffeeF:CoffeeSolveBugsF    4  15.484  3.8709  0.5999 0.66511
## CodingWithoutCoffeeF:GenderF             2  23.715 11.8575  1.8376 0.17382
## CodingWithoutCoffeeF:AgeRangeF           5  28.866  5.7732  0.8947 0.49504
## CoffeeTypeF:CoffeeSolveBugsF            10  77.498  7.7498  1.2010 0.32288
## CoffeeTypeF:GenderF                      4  13.522  3.3806  0.5239 0.71877
## CoffeeTypeF:AgeRangeF                    4  37.839  9.4597  1.4660 0.23282
## CoffeeSolveBugsF:GenderF                 2   4.217  2.1084  0.3267 0.72338
## CoffeeSolveBugsF:AgeRangeF               2   6.947  3.4737  0.5383 0.58835
## GenderF:AgeRangeF                        1   5.911  5.9112  0.9161 0.34489
## Residuals                               36 232.300  6.4528
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Two-way interaction term - Remove CoffeeTypeF - highest p-value (0.42265)
coding5.aov <- aov(ï..CodingHours ~ (CodingWithoutCoffeeF
                   + CoffeeSolveBugsF + GenderF + AgeRangeF)^2,
                   data=cc.data)
anova(coding5.aov)
```

```
## Analysis of Variance Table
##
## Response: ï..CodingHours
##                                         Df Sum Sq Mean Sq F value  Pr(>F)
## CodingWithoutCoffeeF                     2  24.53  12.265  1.8543 0.16417
## CoffeeSolveBugsF                         2  12.49   6.243  0.9439 0.39402
## GenderF                                  1  31.63  31.633  4.7824 0.03210 *
## AgeRangeF                                4  46.95  11.738  1.7746 0.14369
## CodingWithoutCoffeeF:CoffeeSolveBugsF    4   8.33   2.082  0.3148 0.86717
## CodingWithoutCoffeeF:GenderF             2  21.55  10.777  1.6293 0.20342
## CodingWithoutCoffeeF:AgeRangeF           5  40.49   8.098  1.2243 0.30694
## CoffeeSolveBugsF:GenderF                 2   0.45   0.224  0.0339 0.96669
## CoffeeSolveBugsF:AgeRangeF               3   1.67   0.556  0.0840 0.96856
## GenderF:AgeRangeF                        2  36.57  18.283  2.7640 0.06993 .
## Residuals                               70 463.02   6.615
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Two-way interaction term - Remove CoffeeSolvesBugsF - highest p-value (0.39402)
coding6.aov <- aov(ï..CodingHours ~ (CodingWithoutCoffeeF +
                    GenderF + AgeRangeF)^2,
                    data=cc.data)
anova(coding6.aov)
```

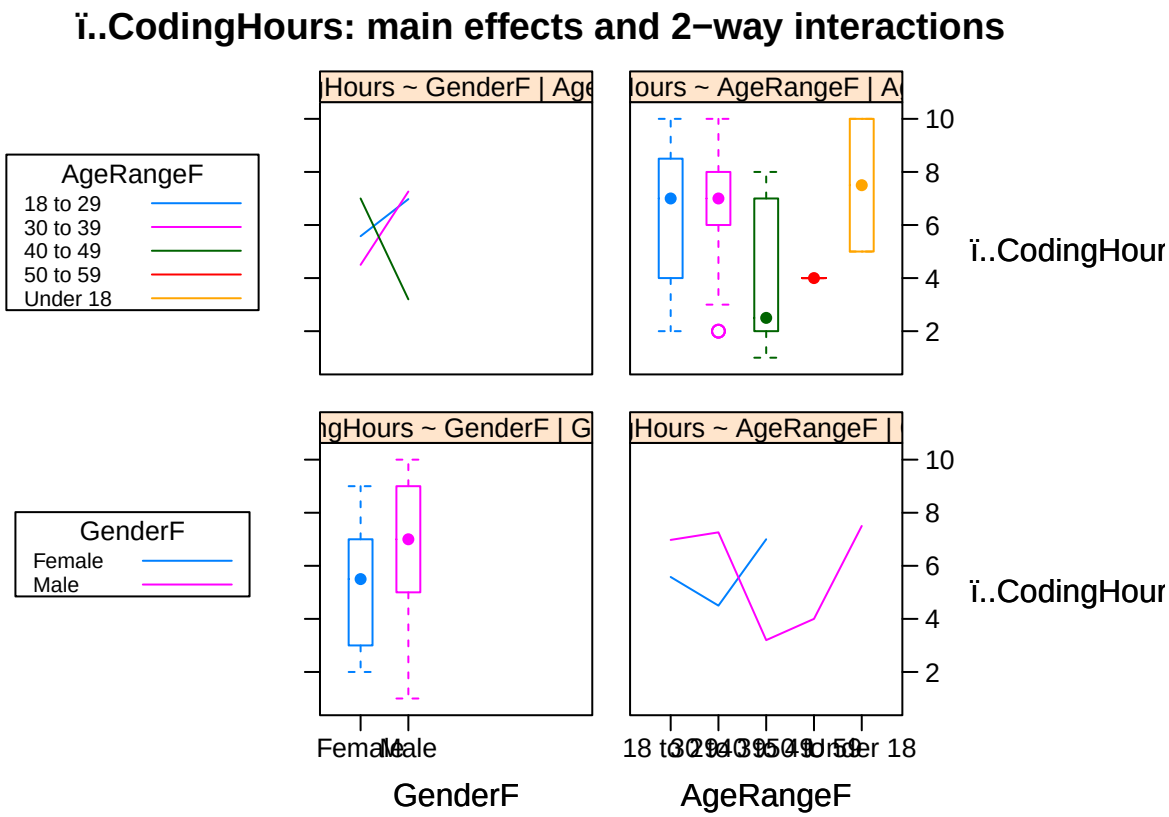
```
## Analysis of Variance Table
##
## Response: ï..CodingHours
##                                Df Sum Sq Mean Sq F value  Pr(>F)
## CodingWithoutCoffeeF            2  24.53  12.265  2.0761 0.13204
## GenderF                         1  33.92  33.921  5.7418 0.01887 *
## AgeRangeF                       4  48.97  12.244  2.0725 0.09200 .
## CodingWithoutCoffeeF:GenderF    2  21.43  10.715  1.8136 0.16962
## CodingWithoutCoffeeF:AgeRangeF  5  47.98   9.596  1.6243 0.16294
## GenderF:AgeRangeF               2  32.30  16.152  2.7340 0.07096 .
## Residuals                      81 478.53   5.908
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Two-way interaction term - Remove CodingWithoutCoffeeF - highest p-value (0.13204)
coding6.aov <- aov(ï..CodingHours ~ (GenderF + AgeRangeF)^2,
                    data=cc.data)
anova(coding6.aov)
```

```
## Analysis of Variance Table
##
## Response: ï..CodingHours
##                   Df Sum Sq Mean Sq F value  Pr(>F)
## GenderF            1  37.08  37.075  5.9284 0.01687 *
## AgeRangeF          4  56.65  14.162  2.2645 0.06835 .
## GenderF:AgeRangeF  2  31.11  15.555  2.4872 0.08884 .
```

```
## Residuals          90 562.84   6.254
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Two-way interaction plot
interaction2wt(ï..CodingHours ~ GenderF + AgeRangeF,
               data=cc.data,
  par.strip.text=list(cex=.8))
```



2. Coffee Cups per Day

```
#cups of coffee drank per day by gender
coffee1.aov <- aov(CoffeeCupsPerDay ~ GenderF, data=cc.data)
anova(coffee1.aov)
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##            Df  Sum Sq Mean Sq F value  Pr(>F)
## GenderF     1  13.539 13.5395  5.4324 0.02182 *
## Residuals 98 244.251  2.4924
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#no interaction term - all factors
coffee2.aov <- aov(CoffeeCupsPerDay ~ CoffeeTimeF + CodingWithoutCoffeeF + CoffeeTypeF
                   + CoffeeSolveBugsF + GenderF + AgeRangeF,
                   data=cc.data)
anova(coffee2.aov)
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                       Df Sum Sq Mean Sq F value    Pr(>F)
## CoffeeTimeF            6 80.578 13.4297 10.3135 2.809e-08 ***
## CodingWithoutCoffeeF   2 18.142  9.0712  6.9664  0.001690 **
## CoffeeTypeF            7 35.985  5.1407  3.9479  0.001031 **
## CoffeeSolveBugsF       2 10.234  5.1168  3.9295  0.023884 *
## GenderF                1  0.007  0.0069  0.0053  0.942367
## AgeRangeF              4  5.210  1.3025  1.0003  0.412997
## Residuals             74 96.359  1.3021
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#2-way interaction term - all factors
coffee3.aov <- aov(CoffeeCupsPerDay ~ (CoffeeTimeF + CodingWithoutCoffeeF + CoffeeTypeF
                   + CoffeeSolveBugsF + GenderF + AgeRangeF)^2,
                   data=cc.data)
anova(coffee3.aov)
```

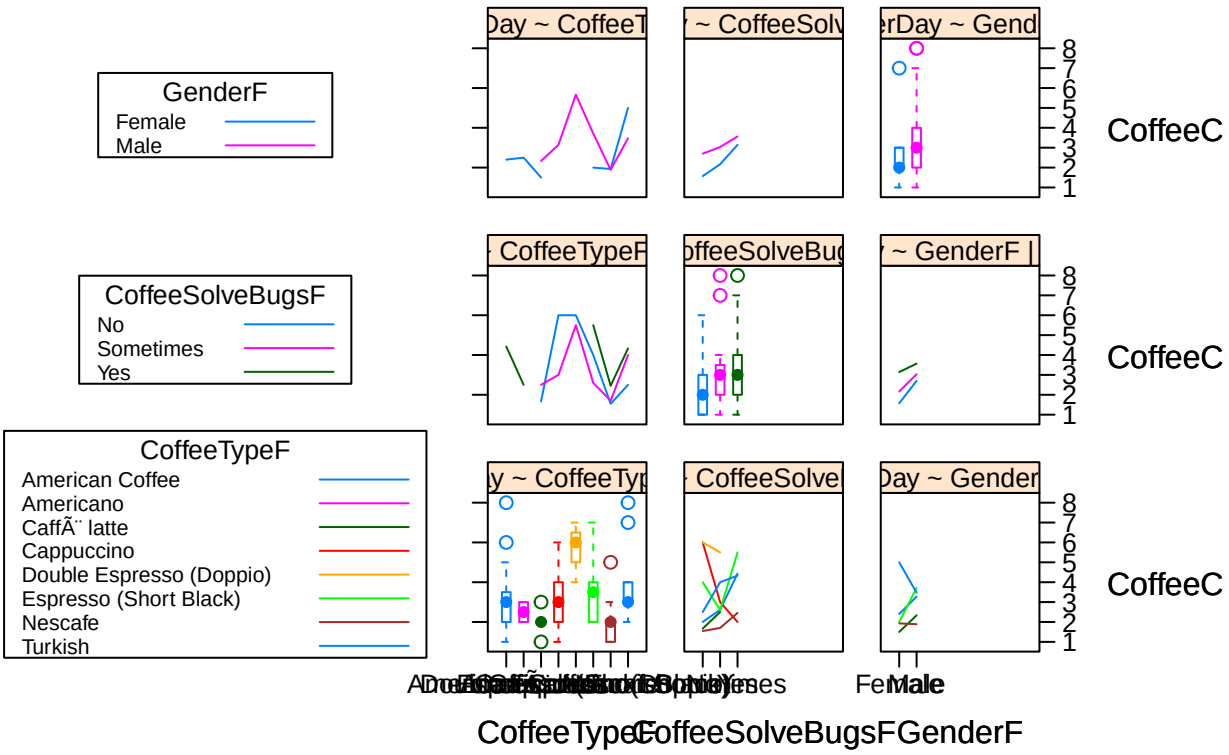
```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                                      Df Sum Sq Mean Sq F value    Pr(>F)
## CoffeeTimeF                           6 80.578 13.4297 11.0737 2.417e-05 ***
## CodingWithoutCoffeeF                  2 18.142  9.0712  7.4798  0.004018 **
## CoffeeTypeF                           7 35.985  5.1407  4.2389  0.005657 **
## CoffeeSolveBugsF                      2 10.234  5.1168  4.2191  0.030463 *
## GenderF                               1  0.007  0.0069  0.0057  0.940867
## AgeRangeF                             4  5.210  1.3025  1.0740  0.396852
## CoffeeTimeF:CodingWithoutCoffeeF      4  8.718  2.1794  1.7971  0.171143
## CoffeeTimeF:CoffeeTypeF              11 21.541  1.9583  1.6147  0.173092
## CoffeeTimeF:CoffeeSolveBugsF          3  0.408  0.1361  0.1122  0.951879
## CoffeeTimeF:GenderF                   1  1.908  1.9077  1.5730  0.224982
## CoffeeTimeF:AgeRangeF                 4  5.307  1.3268  1.0941  0.387797
## CodingWithoutCoffeeF:CoffeeTypeF      6 17.677  2.9462  2.4293  0.065008 .
## CodingWithoutCoffeeF:CoffeeSolveBugsF 4  5.739  1.4347  1.1830  0.349974
## CodingWithoutCoffeeF:GenderF          2  2.372  1.1859  0.9779  0.394254
## CodingWithoutCoffeeF:AgeRangeF        2  1.407  0.7034  0.5800  0.569492
## CoffeeTypeF:CoffeeSolveBugsF          7  3.068  0.4382  0.3614  0.913517
## CoffeeTypeF:GenderF                   4  1.879  0.4697  0.3873  0.815056
## CoffeeTypeF:AgeRangeF                 2  0.568  0.2842  0.2343  0.793370
## CoffeeSolveBugsF:GenderF              2  0.177  0.0883  0.0728  0.930033
## CoffeeSolveBugsF:AgeRangeF            2  2.520  1.2599  1.0388  0.373118
## GenderF:AgeRangeF                     1  0.029  0.0288  0.0237  0.879252
## Residuals                            19 23.042  1.2128
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

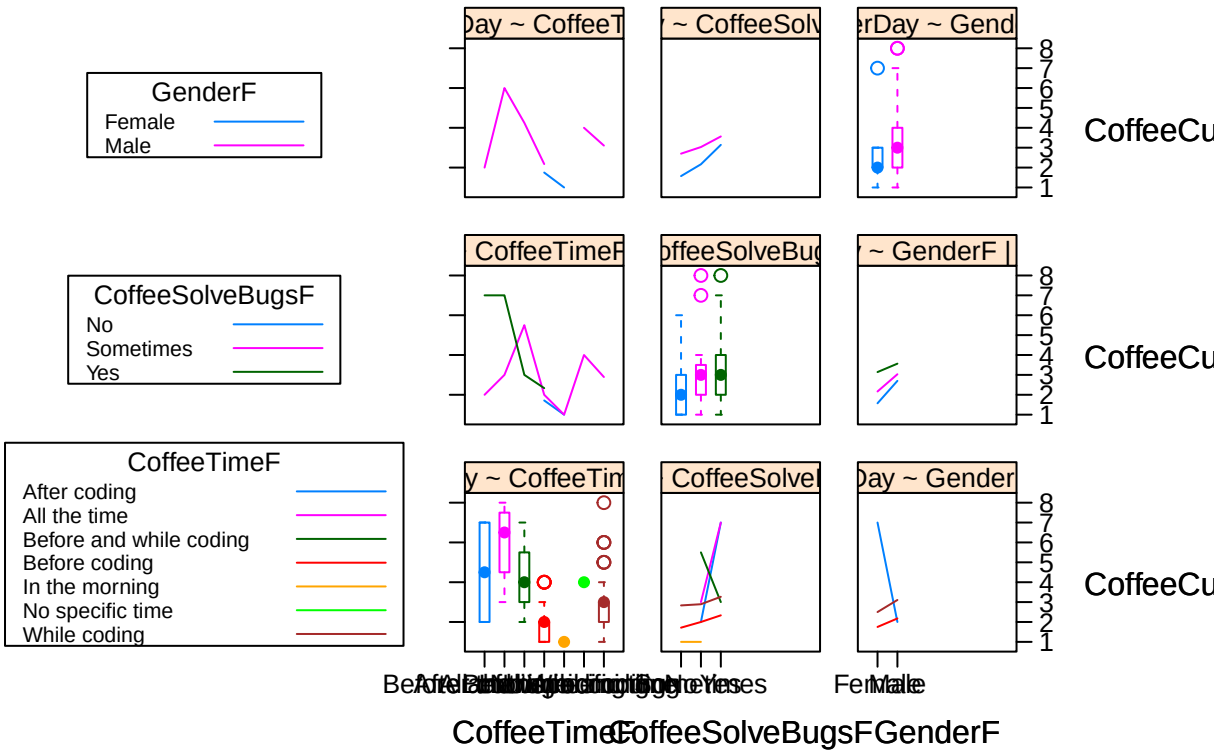
```
#2-way interaction term - gender has the highest p-value 0.940867, remove AgeRangeF first (p-value 0.41
coffee4.aov <- aov(CoffeeCupsPerDay ~ (CoffeeTimeF + CodingWithoutCoffeeF + CoffeeTypeF
                   + CoffeeSolveBugsF + GenderF)^2,
                   data=cc.data)
anova(coffee4.aov)
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                                      Df Sum Sq Mean Sq F value    Pr(>F)
## CoffeeTimeF                           6 81.227 13.5378 14.2031 3.171e-08 ***
## CodingWithoutCoffeeF                  2 17.933  8.9665  9.4071 0.0005168 ***
## CoffeeTypeF                           7 36.754  5.2505  5.5086 0.0002306 ***
## CoffeeSolveBugsF                      2  9.672  4.8362  5.0739 0.0114478 *
## GenderF                               1  0.000  0.0003  0.0003 0.9867520
## CoffeeTimeF:CodingWithoutCoffeeF      4  9.115  2.2786  2.3906 0.0688237 .
## CoffeeTimeF:CoffeeTypeF              11 20.037  1.8216  1.9111 0.0708409 .
## CoffeeTimeF:CoffeeSolveBugsF          3  0.416  0.1386  0.1454 0.9319392
## CoffeeTimeF:GenderF                   1  1.654  1.6545  1.7358 0.1959966
## CodingWithoutCoffeeF:CoffeeTypeF      6  7.732  1.2887  1.3520 0.2602836
## CodingWithoutCoffeeF:CoffeeSolveBugsF 4  7.567  1.8916  1.9846 0.1177204
## CodingWithoutCoffeeF:GenderF          2  2.902  1.4508  1.5221 0.2319697
## CoffeeTypeF:CoffeeSolveBugsF          7 12.366  1.7665  1.8534 0.1068398
## CoffeeTypeF:GenderF                   4  6.245  1.5612  1.6379 0.1859693
## CoffeeSolveBugsF:GenderF              2  0.087  0.0436  0.0458 0.9552947
## Residuals                            36 34.314  0.9532
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

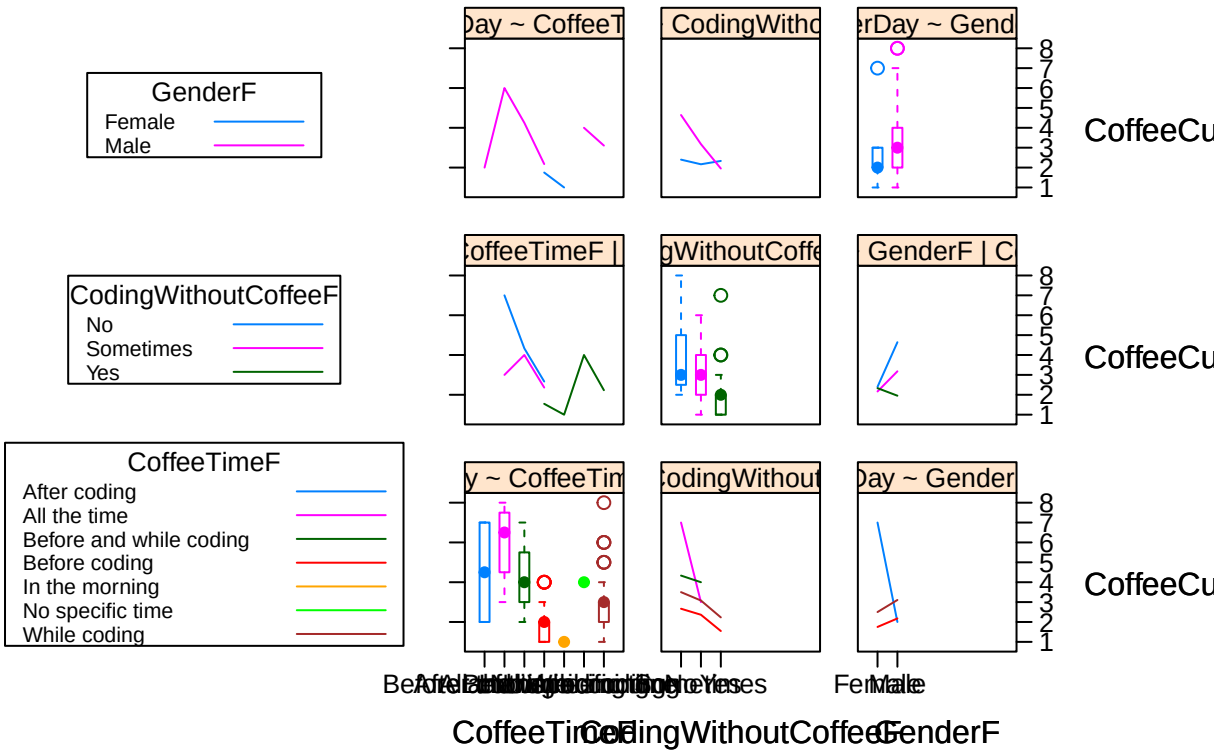
```
#2-way interaction term - gender p-value even higher now (0.9867520), all other main effects are signif

# Two-way interaction plot: CoffeeTypeF, CoffeeSolvesBugsF, GenderF
interaction2wt(CoffeeCupsPerDay ~ CoffeeTypeF + CoffeeSolveBugsF + GenderF,
               data=cc.data,
  par.strip.text=list(cex=.8))
```



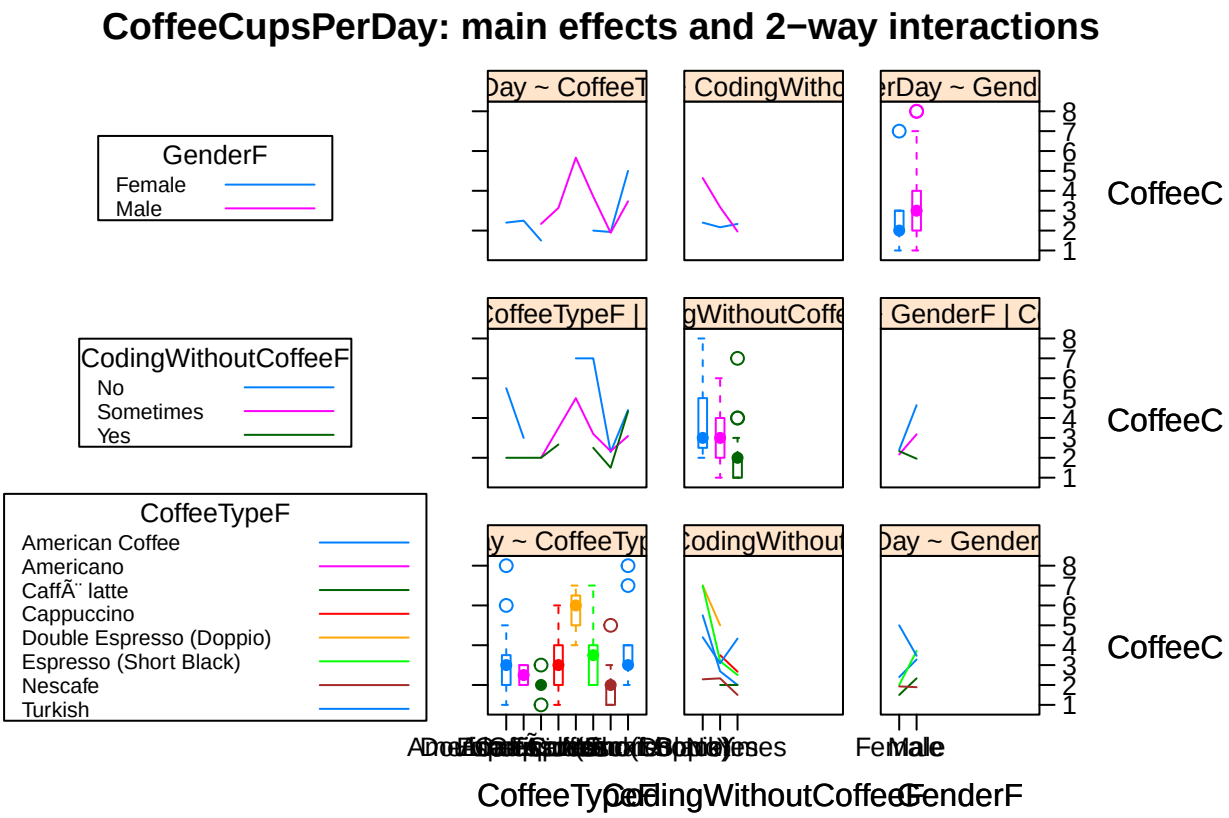

```
# Two-way interaction plot: CoffeeTimeF, CoffeeSolvesBugsF, GenderF
interaction2wt(CoffeeCupsPerDay ~ CoffeeTimeF + CoffeeSolveBugsF + GenderF,
               data=cc.data,
  par.strip.text=list(cex=.8))
```




```
# Two-way interaction plot: CoffeeTimeF, CodingWithoutCoffee, GenderF
interaction2wt(CoffeeCupsPerDay ~ CoffeeTimeF + CodingWithoutCoffeeF + GenderF,
               data=cc.data,
  par.strip.text=list(cex=.8))
```




```
# Two-way interaction plot: CoffeeTypeF, CodingWithoutCoffee, GenderF
interaction2wt(CoffeeCupsPerDay ~ CoffeeTypeF + CodingWithoutCoffeeF + GenderF,
               data=cc.data,
  par.strip.text=list(cex=.8))
```



#Part 2 Findings One-way anovas between gender and coding hours and gender and coffee cups per day both show that gender is statistically significant at p-values of 0.02075 and 0.02182 respectively.

Looking at an anova including all factors and two-way interactions, gender is the only statistically significant predictor for coding hours at a p-value of 0.04443 and gender's interaction with age range is significant at a 0.1 alpha but not at 0.05, with a p-value of 0.9688. We went through a stepwise process to trim the model down, removing the factors with the highest p-values one a time, which resulted in a two-way anova modeling coding hours by gender and age range. Here, gender is still the most statistically significant variable with a p-value of 0.01687, and age range and the interaction between gender and age range are both only significant at a 0.1 alpha. Looking at a two-way interaction plot between gender and age as it relates to coding hours shows that the interquartile ranges overlap for gender, but there is more variability for males (and more data points). The median for males is above the 3rd quartile for females. For age range, the 40-49 year-old group has a lower median than the other groups, but all interquartile ranges overlap (excluding the 50-59 year-old group which has limited data), so it makes sense that the p-value for age range was only significant at a 0.1 alpha. When looking at coding hours, we disagree with the statement that gender is not significant when you consider the other given variables. Gender is the most significant predictor for coding hours in this data set.

However, we agree that gender is not statistically significant for coffee cups per day when considering other predictors. With all factors and two-way interactions in an anova for coffee cups per day, gender's p-value is high at 0.940867 and no interactions with gender are significant. We started to follow a similar process to trim the model by removing the variables with the highest p-values. Gender's p-value of 0.940867 is highest, so we first removed age range, which has a p-value 0.412997. This made the p-value for gender even higher at 0.9867520 and all remaining variables have significant p-values for main effects at at least a 0.05 alpha, but no interactions are significant at a 0.05 alpha with this many predictors included. Since gender has the highest p-value when considering other variables, while the other variables are statistically significant with low p-values, we agree that when you consider other variables, such as coffee time, coding without coffee, coffee type, and coffee solves bugs, differences in these variables explain the differences in coffee cups

per day and gender is not significant. We included four two-way interaction plots, showing gender coupled with combinations of two at a time from the four variables that showed up in the anova as being significant so we could see their interactions. The main effects graphs indicate variation among the levels for Coffee Time, Coffee Type, and Coding Without Coffee, but don't appear to show much variation for Coffee Solve Bugs or Gender, where all interquartile ranges overlap. Coffee Time shows higher consumption for After Coding, All the Time, and Before and While Coding, which overlap each other, and lower consumption for Before Coding and While Coding. In the morning is lowest, but with limited observations. There are also limited observations for No Specific Time. Coffee Type shows lower consumption for Nescafe and higher consumption for Double Espresso, while the other coffee types all overlap. Coding Without Coffee shows lower consumption for "Yes," while "No" and "Sometimes" overlap. With the exception of Gender and Coffee Solve Bugs, all other interaction plots show evidence of interaction.

Part 3

Design a two-way analysis of variance model to model Coffee Cups Per Day. Your model should have all significant terms – two main effects and interaction. Explain your model in detail. How did you select your two factors? What factor combinations have increased or decreased coffee cups per day? Include an interaction plot and MMC analysis output. Both should be explained.

Does adding a coding hours per day as a co-variate, improve your model? Explain.

See Helpful Notes in the assignment document.

Check each combination of 2 factors to compare options

```
#factors are significant, interaction is not
coffee3a.aov <- aov(CoffeeCupsPerDay ~ (CoffeeTimeF + CodingWithoutCoffeeF)^2,
                    data=cc.data)
anova(coffee3a.aov)
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                                  Df  Sum Sq Mean Sq F value   Pr(>F)
## CoffeeTimeF                       6  81.646 13.6076  7.9626 7.43e-07 ***
## CodingWithoutCoffeeF              2  18.275  9.1376  5.3470 0.006454 **
## CoffeeTimeF:CodingWithoutCoffeeF  4   9.192  2.2981  1.3448 0.259884
## Residuals                        87 148.677  1.7089
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#All significant with lowest Sum Sq - 89.632 Sum Sq
coffee3b.aov <- aov(CoffeeCupsPerDay ~ (CoffeeTimeF + CoffeeTypeF)^2,
                    data=cc.data)
anova(coffee3b.aov)
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                          Df Sum Sq Mean Sq F value    Pr(>F)
## CoffeeTimeF               6 81.227 13.5378 11.0258 1.027e-08 ***
```

```
## CoffeeTypeF              7 40.613  5.8018  4.7252 0.0002043 ***
## CoffeeTimeF:CoffeeTypeF 12 36.549  3.0457  2.4806 0.0088132 **
## Residuals               73 89.632  1.2278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#CoffeeSolveBugsF is not significant
coffee3c.aov <- aov(CoffeeCupsPerDay ~ (CoffeeTimeF + CoffeeSolveBugsF)^2,
                    data=cc.data)
anova(coffee3c.aov)
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                              Df  Sum Sq Mean Sq F value    Pr(>F)
## CoffeeTimeF                   6  81.646 13.6076  8.1431 5.794e-07 ***
## CoffeeSolveBugsF              2   5.382  2.6910  1.6103   0.20586
## CoffeeTimeF:CoffeeSolveBugsF  6  28.722  4.7870  2.8646   0.01368 *
## Residuals                    85 142.040  1.6711
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#All significant - 115.046 Sum Sq
coffee3d.aov <- aov(CoffeeCupsPerDay ~ (CoffeeTypeF + CoffeeSolveBugsF)^2,
                    data=cc.data)
anova(coffee3d.aov)
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                              Df  Sum Sq Mean Sq F value    Pr(>F)
## CoffeeTypeF                   7  73.031 10.4329  7.1641  1.26e-06 ***
## CoffeeSolveBugsF              2  24.597 12.2985  8.4452 0.0004744 ***
## CoffeeTypeF:CoffeeSolveBugsF 10  35.347  3.5347  2.4272 0.0141152 *
## Residuals                    79 115.046  1.4563
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#All significant - 113.608 Sum Sq
coffee3e.aov <- aov(CoffeeCupsPerDay ~ (CoffeeTypeF + CodingWithoutCoffeeF)^2,
                    data=cc.data)
anova(coffee3e.aov)
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                                  Df  Sum Sq Mean Sq F value    Pr(>F)
## CoffeeTypeF                       7  73.031 10.4329  7.2548 1.057e-06 ***
## CodingWithoutCoffeeF              2  32.536 16.2681 11.3124 4.784e-05 ***
## CoffeeTypeF:CodingWithoutCoffeeF 10  28.845  2.8845  2.0058   0.04354 *
## Residuals                        79 113.608  1.4381
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#CoffeeSolveBugsF is not significant
coffee3f.aov <- aov(CoffeeCupsPerDay ~ (CodingWithoutCoffeeF + CoffeeSolveBugsF)^2,
                    data=cc.data)
anova(coffee3f.aov)
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                                      Df  Sum Sq Mean Sq F value    Pr(>F)
## CodingWithoutCoffeeF                  2  46.152 23.0762 11.3668 3.922e-05 ***
## CoffeeSolveBugsF                      2   6.564  3.2821  1.6167   0.20421
## CodingWithoutCoffeeF:CoffeeSolveBugsF  4  20.330  5.0825  2.5035   0.04767 *
## Residuals                             91 184.743  2.0301
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#All significant, 182.8 Sum Sq
coffee3g.aov <- aov(CoffeeCupsPerDay ~ (GenderF + CodingWithoutCoffeeF)^2,
                    data=cc.data)
anova(coffee3g.aov)
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                              Df  Sum Sq Mean Sq F value    Pr(>F)
## GenderF                       1  13.539 13.5395  6.9632  0.009739 **
## CodingWithoutCoffeeF          2  44.603 22.3016 11.4695 3.491e-05 ***
## GenderF:CodingWithoutCoffeeF  2  16.870  8.4352  4.3381  0.015774 *
## Residuals                    94 182.777  1.9444
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#interaction not significant
coffee3h.aov <- aov(CoffeeCupsPerDay ~ (GenderF + CoffeeSolveBugsF)^2,
                    data=cc.data)
anova(coffee3h.aov)
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                           Df  Sum Sq Mean Sq F value  Pr(>F)
## GenderF                    1  13.539 13.5395  5.6052 0.01996 *
## CoffeeSolveBugsF           2  15.833  7.9166  3.2774 0.04208 *
## GenderF:CoffeeSolveBugsF   2   1.359  0.6796  0.2814 0.75539
## Residuals                 94 227.058  2.4155
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#Age range not significant
coffee3i.aov <- aov(CoffeeCupsPerDay ~ (GenderF + AgeRangeF)^2,
                    data=cc.data)
anova(coffee3i.aov)
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                   Df  Sum Sq Mean Sq F value   Pr(>F)
## GenderF            1  14.447 14.4470  6.2394 0.014311 *
## AgeRangeF          4   5.684  1.4210  0.6137 0.653850
## GenderF:AgeRangeF  2  27.651 13.8255  5.9710 0.003673 **
## Residuals         90 208.391  2.3155
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Age range & interaction are not significant
coffee3j.aov <- aov(CoffeeCupsPerDay ~ (AgeRangeF + CodingWithoutCoffeeF)^2,
                    data=cc.data)
anova(coffee3j.aov)
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                                Df  Sum Sq Mean Sq F value   Pr(>F)
## AgeRangeF                       4   3.805  0.9513  0.4172   0.7958
## CodingWithoutCoffeeF            2  47.118 23.5590 10.3321 9.52e-05 ***
## AgeRangeF:CodingWithoutCoffeeF  5   9.154  1.8309  0.8030   0.5506
## Residuals                      86 196.096  2.2802
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Age range & interaction are not significant
coffee3k.aov <- aov(CoffeeCupsPerDay ~ (AgeRangeF + CoffeeSolveBugsF)^2,
                    data=cc.data)
anova(coffee3k.aov)
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                             Df  Sum Sq Mean Sq F value  Pr(>F)
## AgeRangeF                    4   3.805  0.9513  0.3520 0.84195
## CoffeeSolveBugsF             2  17.476  8.7380  3.2337 0.04425 *
## AgeRangeF:CoffeeSolveBugsF   5   2.507  0.5013  0.1855 0.96736
## Residuals                   86 232.386  2.7022
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#Selected two-way anova

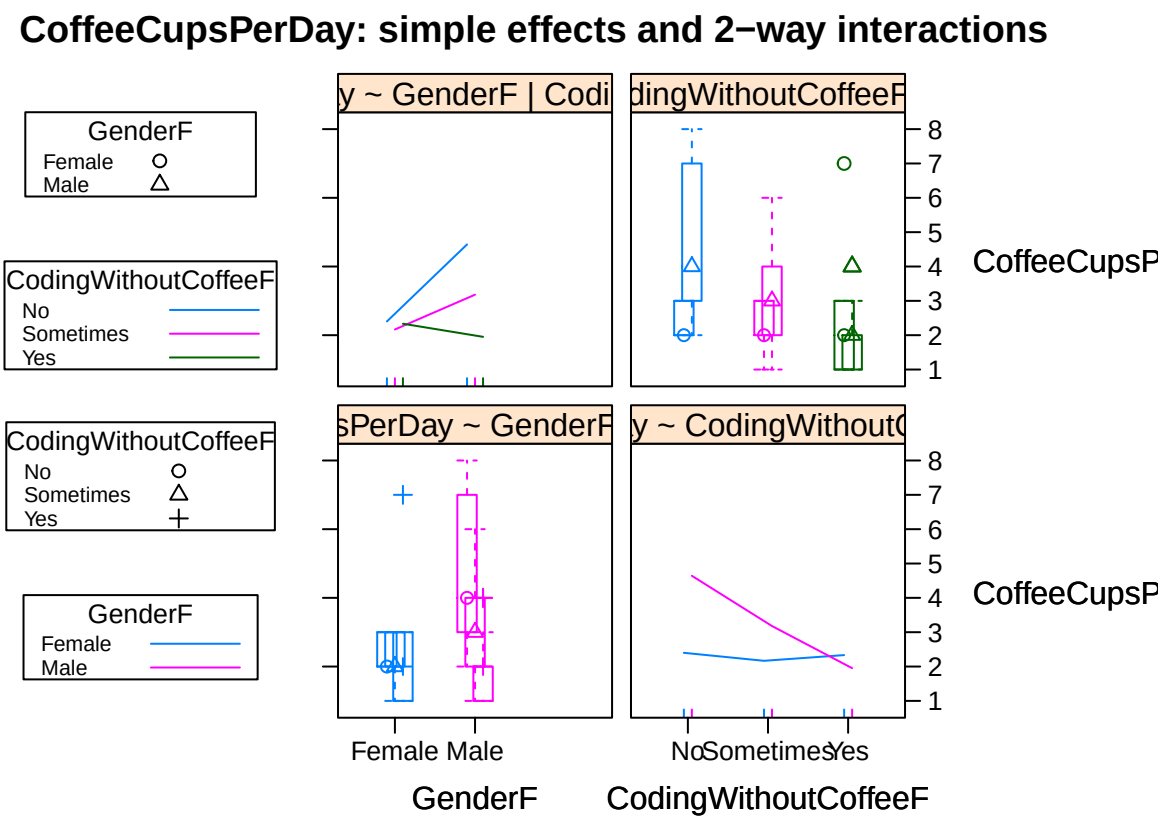
```
coffee3g.aov <- aov(CoffeeCupsPerDay ~ (GenderF + CodingWithoutCoffeeF)^2,
                    data=cc.data)
anova(coffee3g.aov)
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
```

```
##                                Df  Sum Sq Mean Sq F value    Pr(>F)
## GenderF                         1  13.539 13.5395  6.9632  0.009739 **
## CodingWithoutCoffeeF            2  44.603 22.3016 11.4695 3.491e-05 ***
## GenderF:CodingWithoutCoffeeF    2  16.870  8.4352  4.3381  0.015774 *
## Residuals                      94 182.777  1.9444
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#Interaction Plot displaying simple effects

```
interaction2wt(data=cc.data, CoffeeCupsPerDay ~ GenderF + CodingWithoutCoffeeF,
                 simple=TRUE,
               between=list(x=.5, y=.5))
```



#Mean Minus Mean Comparison (MMC) Analysis #3g - Gender & Coding Without Coffee

```
# separate ANOVA for each gender
coffee.aov.2g <- sapply(levels(
  cc.data$GenderF),
  function(i) aov(CoffeeCupsPerDay ~ CodingWithoutCoffeeF, data = cc.data,
      subset=(GenderF==i)),
      simplify=FALSE)

print(lapply(coffee.aov.2g, anova))
```

```
## $Female
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                       Df Sum Sq Mean Sq F value Pr(>F)
## CodingWithoutCoffeeF  2  0.249 0.12436  0.0776 0.9256
## Residuals            23 36.867 1.60290
##
## $Male
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                       Df  Sum Sq Mean Sq F value    Pr(>F)
## CodingWithoutCoffeeF  2  61.225 30.6124  14.896 3.963e-06 ***
## Residuals            71 145.910  2.0551
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Adjustments needed for Multiple Comparisons
ResidMS <- function(x)
  summary(x)[[1]]["Residuals","Mean Sq"]

ResidMSAvg <- ResidMS(coffee3g.aov)
ResidMSAvg
```

```
## [1] 1.944435
```

```
crit.val <- qtukey(p = 0.95, nmeans = 3, df = 94)/sqrt(2)
crit.val
```

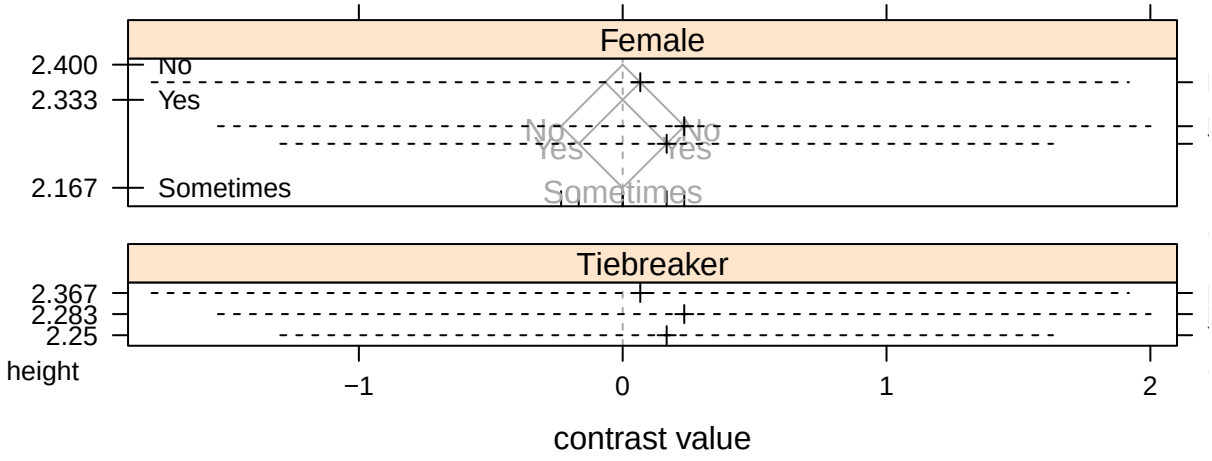
```
## [1] 2.381402
```

```
# Multiple Comparison Results by Gender
coffee.mmc.2g <- sapply(
  coffee.aov.2g, simplify = FALSE,
  function(x) mmc(x,
                  calpha = crit.val
                  * sqrt(ResidMSAvg/ResidMS(x))))
coffee.mmc.2g
```

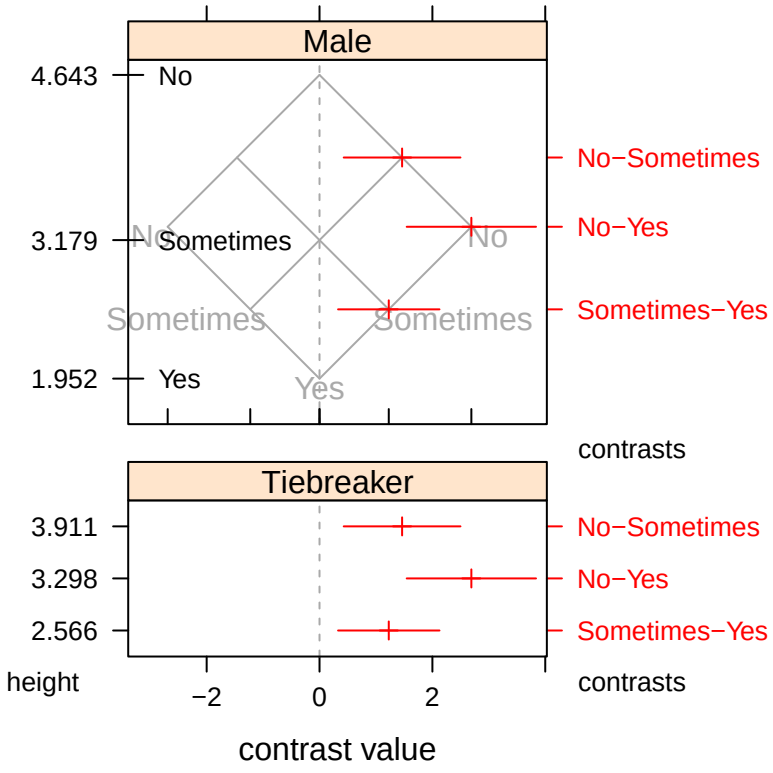
```
## $Female
## Tukey contrasts
## Fit: aov(formula = CoffeeCupsPerDay ~ CodingWithoutCoffeeF, data = cc.data, subset = (GenderF ==
##           i))
## Estimated Quantile = 2.622868
## 95% family-wise confidence level
## $mca
##                   estimate    stderr     lower    upper   height
## No-Yes         0.06666667 0.7061725 -1.785531 1.918864 2.366667
## No-Sometimes   0.23333333 0.6739099 -1.534244 2.000910 2.283333
## Yes-Sometimes 0.16666667 0.5582784 -1.297624 1.630957 2.250000
## $none
##           estimate    stderr     lower    upper   height
```

```
## No        2.400000 0.5661976 0.9149382 3.885062 2.400000
## Yes       2.333333 0.4220188 1.2264336 3.440233 2.333333
## Sometimes 2.166667 0.3654790 1.2080634 3.125270 2.166667
##
## $Male
## Tukey contrasts
## Fit: aov(formula = CoffeeCupsPerDay ~ CodingWithoutCoffeeF, data = cc.data, subset = (GenderF ==
##           i))
## Estimated Quantile = 2.316412
## 95% family-wise confidence level
## $mca
##               estimate    stderr     lower    upper   height
## No-Sometimes  1.463370 0.4466377 0.4287732 2.497967 3.911172
## No-Yes        2.690476 0.4946228 1.5447262 3.836226 3.297619
## Sometimes-Yes 1.227106 0.3880140 0.3283060 2.125906 2.565934
## $none
##           estimate    stderr    lower    upper   height
## No        4.642857 0.3831331 3.755363 5.530351 4.642857
## Sometimes 3.179487 0.2295522 2.647750 3.711225 3.179487
## Yes       1.952381 0.3128269 1.227745 2.677017 1.952381
```

```r
# coffee.mmc.2g graphic
mmc2g <- sapply(coffee.mmc.2g,
                mmcplot, style="both",
                simplify=FALSE, axis.right=2,
                ylab.right=NULL, ylab=NULL)

# mmc2g
mmc2g[[1]]$condlevels[[1]][1] <-
  names(mmc2g)[1]
mmc2g[[2]]$condlevels[[1]][1] <-
  names(mmc2g)[2]
old.digits <- options(digits=4)

# Prints each level of Gender individually
print(mmc2g[[1]])
```




```
print(mmc2g[[2]])
```



Add covariate to selected two-way anova

```
coffee3g.cov.aov <- aov(CoffeeCupsPerDay ~ (GenderF + CodingWithoutCoffeeF)^2 + ï..CodingHours,
                    data=cc.data)
anova(coffee3g.cov.aov)
```

```
## Analysis of Variance Table
##
## Response: CoffeeCupsPerDay
##                               Df Sum Sq Mean Sq F value  Pr(>F)
## GenderF                        1   13.5   13.54    7.13   0.009 **
## CodingWithoutCoffeeF           2   44.6   22.30   11.74 2.8e-05 ***
## ï..CodingHours                 1   10.0   10.03    5.28   0.024 *
## GenderF:CodingWithoutCoffeeF   2   13.0    6.49    3.41   0.037 *
## Residuals                     93  176.6    1.90
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Partial f-test to two-way anova with and without covariate

```
anova(coffee3g.aov, coffee3g.cov.aov)
```

```
## Analysis of Variance Table
##
## Model 1: CoffeeCupsPerDay ~ (GenderF + CodingWithoutCoffeeF)^2
## Model 2: CoffeeCupsPerDay ~ (GenderF + CodingWithoutCoffeeF)^2 + ï..CodingHours
##   Res.Df RSS Df Sum of Sq    F Pr(>F)
## 1     94 183
## 2     93 177  1      6.13 3.23  0.076 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Part 3 Explanation and Findings

To select our factors for a two-way analysis of variance, we considered each possible combination of two factors to see which had significant main effects and interactions. This resulted in four potential options with significance: Coffee Type and Coffee Time, Coffee Type and Coffee Solve Bugs, Coffee Type and Coding Without Coffee, and Gender and Coding Without Coffee. Of these options, Coffee Type and Coffee Time resulted in the lowest residual sum of squares at 89.632; however, there are 8 levels of Coffee Type and 7 levels of Coffee Time and each have very limited observations within certain levels. We believe this model had the lowest residual sum of squares because it is overfitting the data, since there are several combinations with only 1 observation. To a lesser extent, the same problem exists in any model that uses Coffee Type or Coffee Time with another factor. Even though our two-way anova containing Gender and Coding Without Coffee has the highest residual sum of squares of these four options at 182.8, it avoids overfitting and does not require removing observations or manipulating the data to continue the analysis. This is because there are only 6 combinations of gender and coding without coffee compared to 56 combinations of Coffee Time and Coffee Type and 24 combinations of Coffee Type with either Coffee Solve Bugs or Coding Without Coffee, many of which have less than 2 observations.For these reasons, we selected the two-way anova modeling coffee cups per day by Gender and Coding Without Coffee.

The two-way interaction plot between Gender and Coding Without Coffee with simple effects appears to indicate that there are not significant differences in coffee consumption between coding without coffee for females but there are for males. The lines on the interaction plots cross and show higher coffee consumption for females than males for the group of coders who do not drink coffee while coding (i.e., Coding Without Coffee = Yes), while males have higher mean coffee consumption than females for the other levels. The bottom left graph of simple effects shows that the levels of coding without coffee for females are clumped together, while the levels for males show variation. The levels Yes and No do not overlap. The bottom of the interquartile range for No overlaps with the top of the interquartile range for Sometimes. The top right graph of simple effects shows that there is no overlap in male and female interquartile ranges for No, while the interquartile ranges overlap for Sometimes and Yes.

The mean minus mean comparison confirms what we observed in the two-way interaction plot. It shows that there is no significant difference in the levels of Coding Without Coffee for Females, but there is a significant difference in all levels of Coding Without Coffee for Males. With Males, the most coffee is consumed when coding with coffee (i.e., Coding Without Coffee = No), followed by sometimes coding with coffee, and the least coffee is consumed when coding without coffee.

Next we added coding hours to the model as a covariate, marginally improving the residual sum of squares from 182.8 to 176.6. A partial f-test comparing the models had a p-value of 0.076, indicating that the improvement is only statistically signficant at a 0.1 alpha and not at 0.05.