Assignment 2

Alex Teboul

Due: 2/13/2019 by 11:59pm

Class

DSC 423 - Data Analysis and Regression
Total points: 43pts

Problem 1 [28 points]

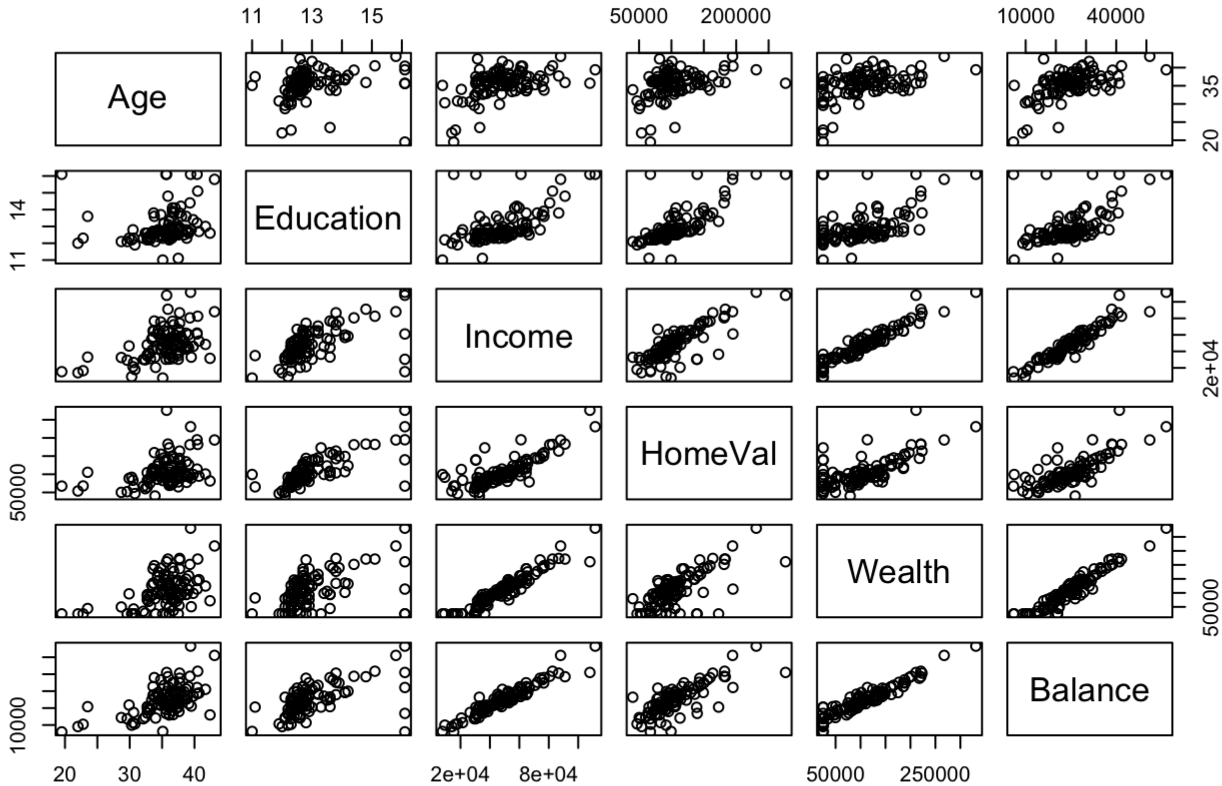
The file bankingfull.txt attached to this assignment contains the full dataset. You analyzed a smaller set for Assignment 1. It provides data acquired from banking and census records for different zip codes in the bank's current market. Such information can be useful in targeting advertising for new customers or for choosing locations for branch offices. The data show: * median age of the population (AGE) * median years of education (EDUCATION) * median income (INCOME) in $ * median home value (HOMEVAL) in $ * median household wealth (WEALTH) in $ * average bank balance (BALANCE) in $ **The goal of this exercise is to define a regression model to predict the average bank balance as a function of the other variables.**

Problem 1 a)

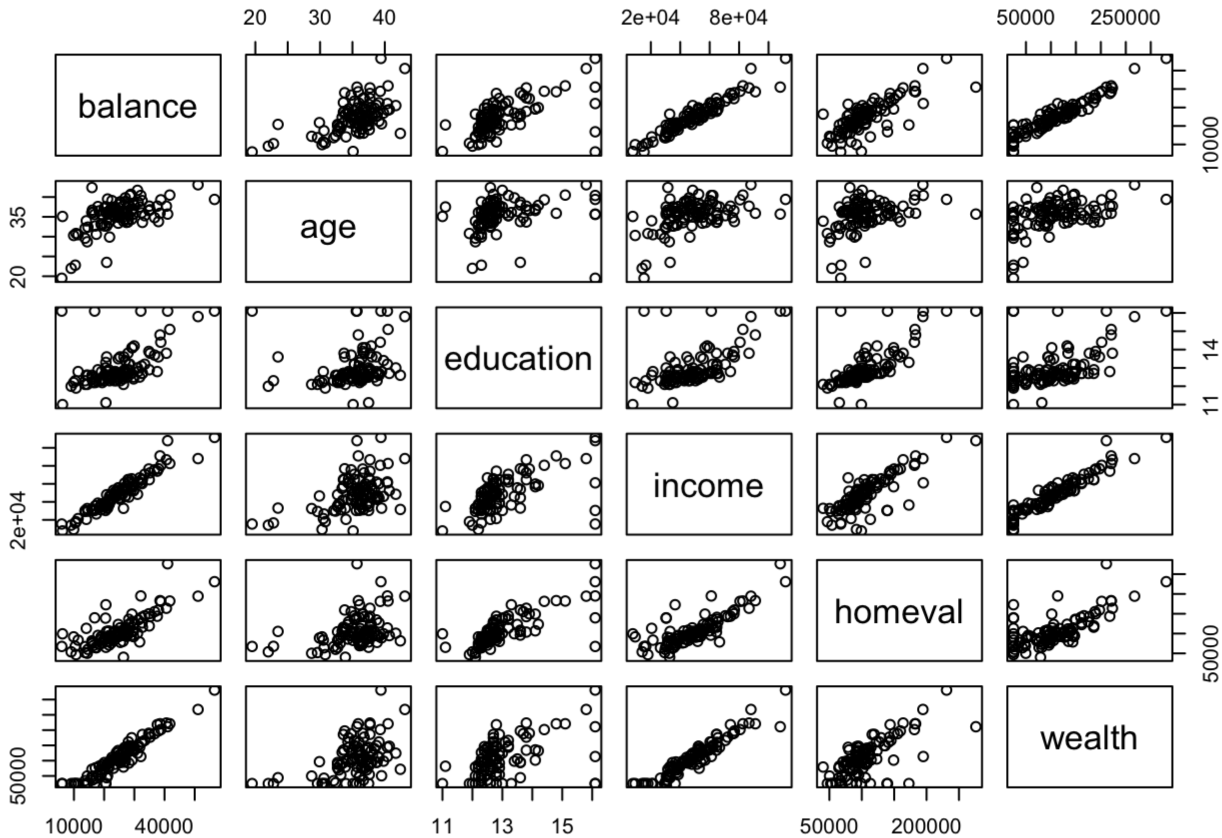
Create scatterplots to visualize the associations between bank balance and the other five variables. Discuss the patterns displayed by the scatterplot. Do the associations appear to be linear? (you can create scatterplots or a matrix plot) [1 pt R code, 1 pt scatterplots, 2 pts answer = 4 pts]

```
## load in the data from file
myd=read.table("Bankingfull.txt", header=T, sep="\t")

## get the variables
balance=myd$Balance

age=myd$Age
education=myd$Education
income=myd$Income
homeval=myd$HomeVal
wealth=myd$Wealth

# general scatterplot matrix for quantitative variables
plot(myd, main="Simple Scatterplot Matrix")
```

Simple Scatterplot Matrix



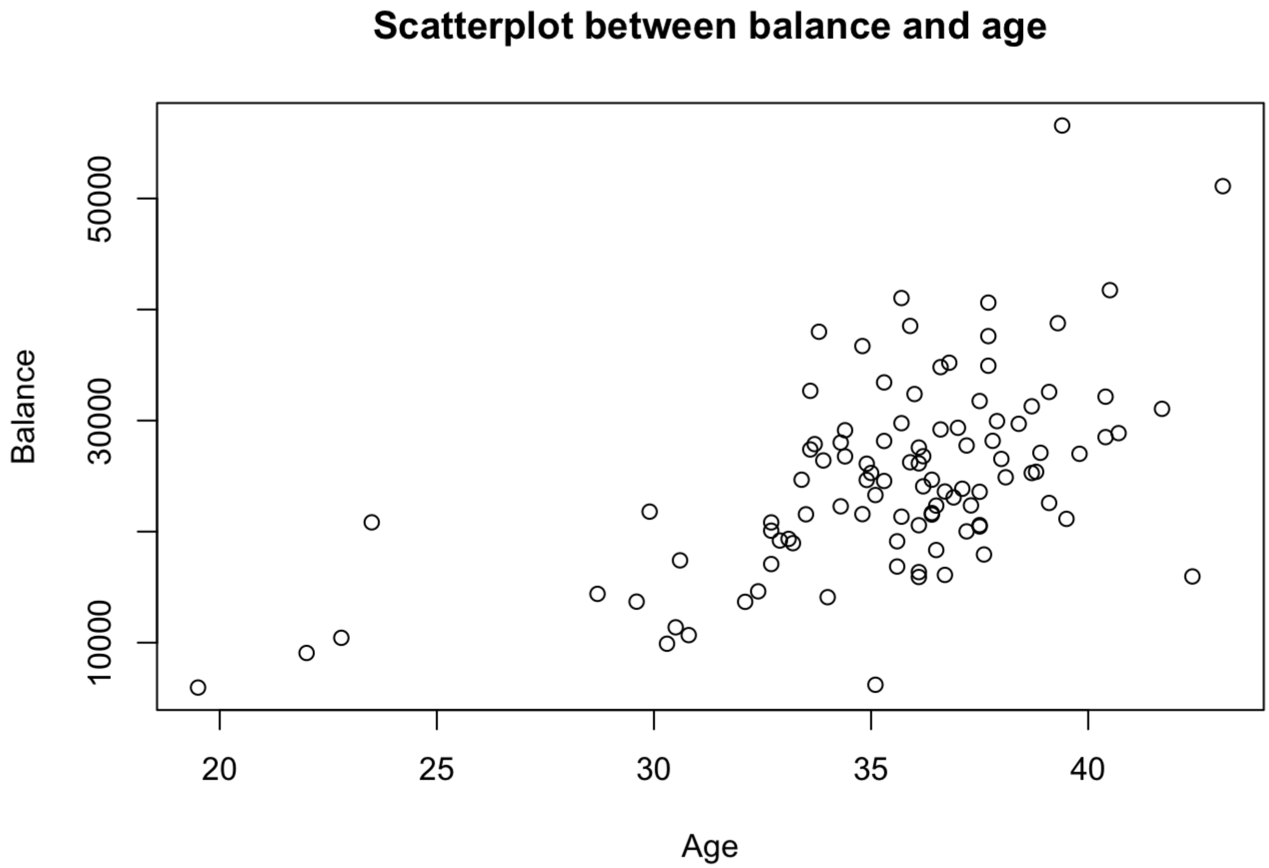

```
# specific pairs plot balance
pairs(balance ~ age+education+income+homeval+wealth)
```



Answer:

Balance appears to be positively associated with all the variables. Some of the relationships do appear to be linear. I will look at the individual plots below for for specificity.

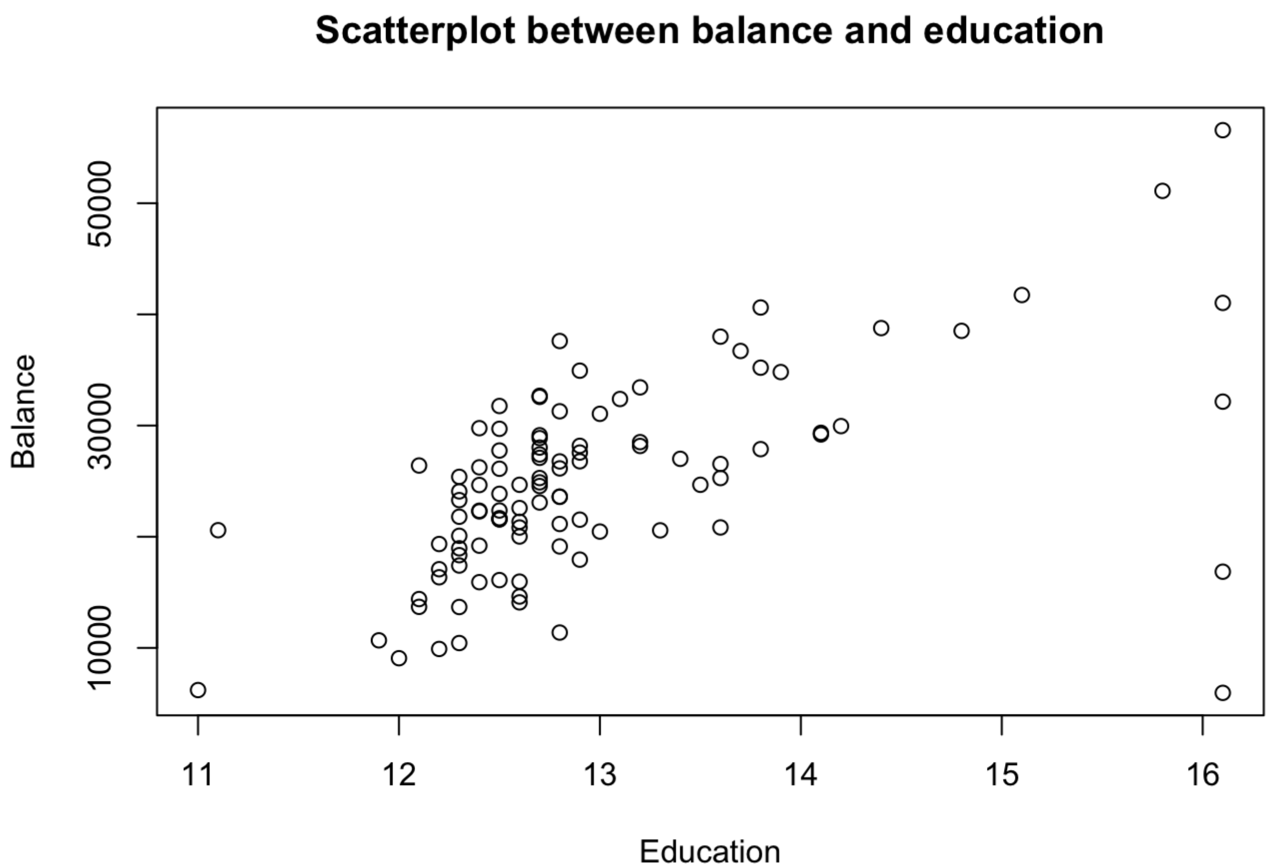
```
# 1 scatterplot between balance and age
plot(age, balance, main="Scatterplot between balance and age", xlab="Age", ylab="Balanc
e")
```



Answer:

The relationship between balance and age appears to be somewhat linear. Not great though.

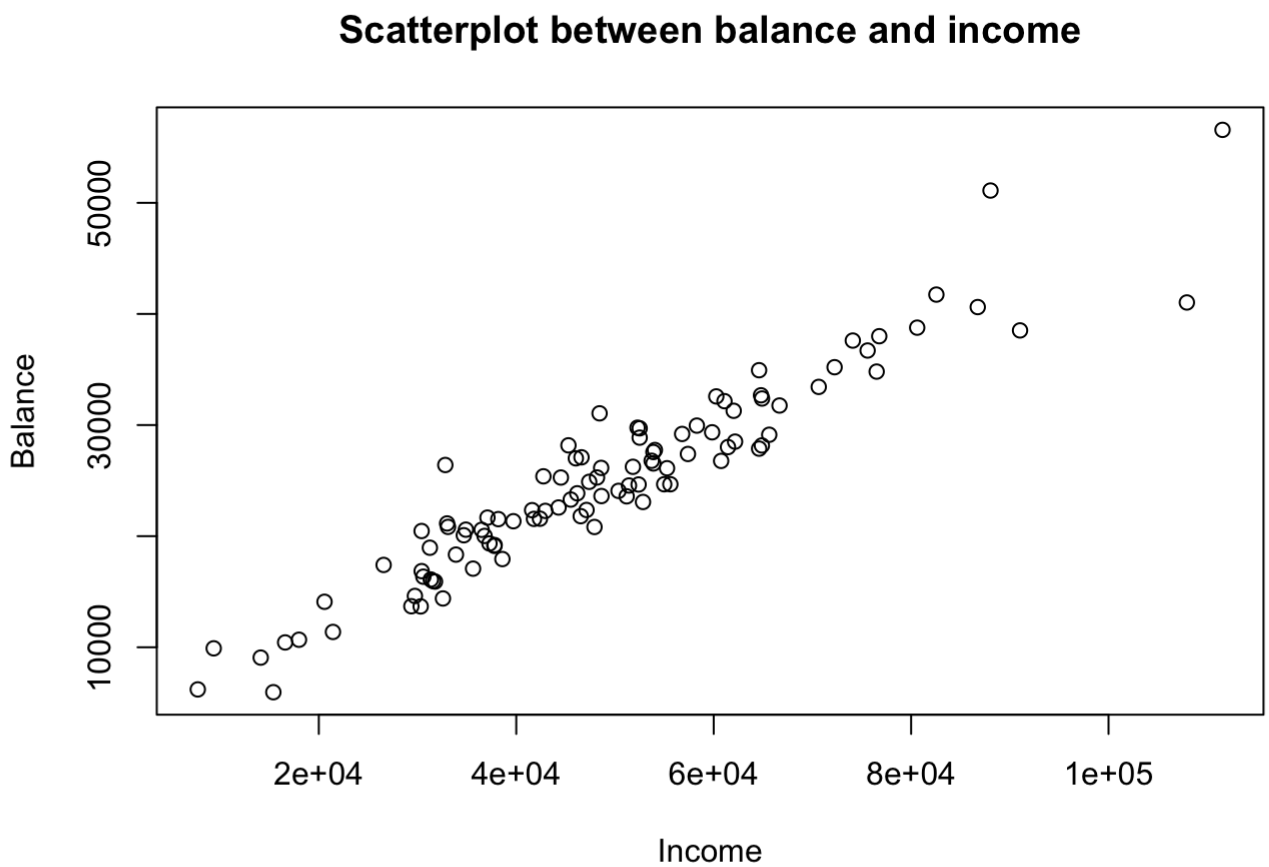
```
# 2 scatterplot between balance and education
plot(education, balance, main="Scatterplot between balance and education", xlab="Educati
on", ylab="Balance")
```



Answer:

The relationship between balance and education appears to be somewhat linear as well. That said, the education values above 16 do not follow the trend and would throw off a line of best fit.

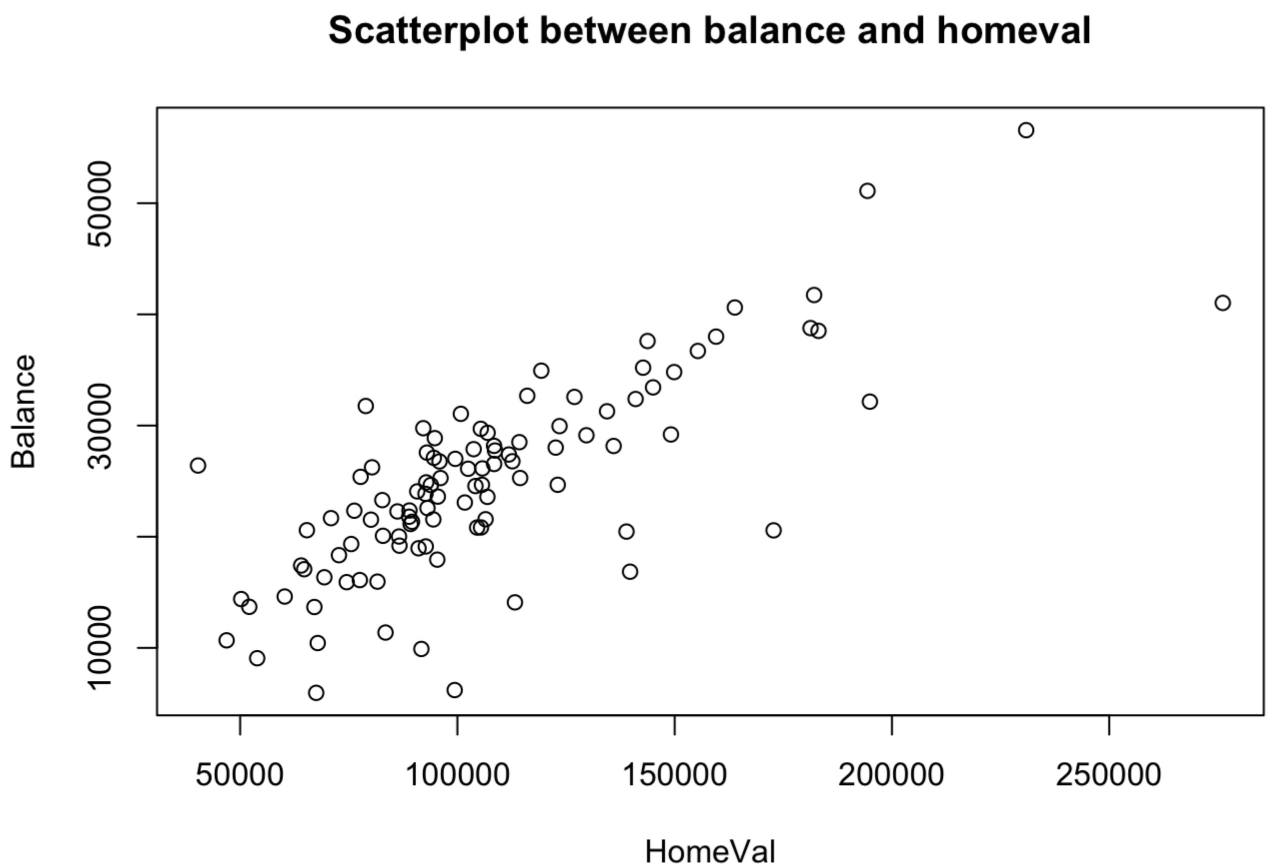
```
# 3 scatterplot between balance and income
plot(income, balance, main="Scatterplot between balance and income", xlab="Income", ylab
="Balance")
```



Answer:

The relationship between balance and income can be described as linear. The relationship is more clearly linear than for age or eduation.

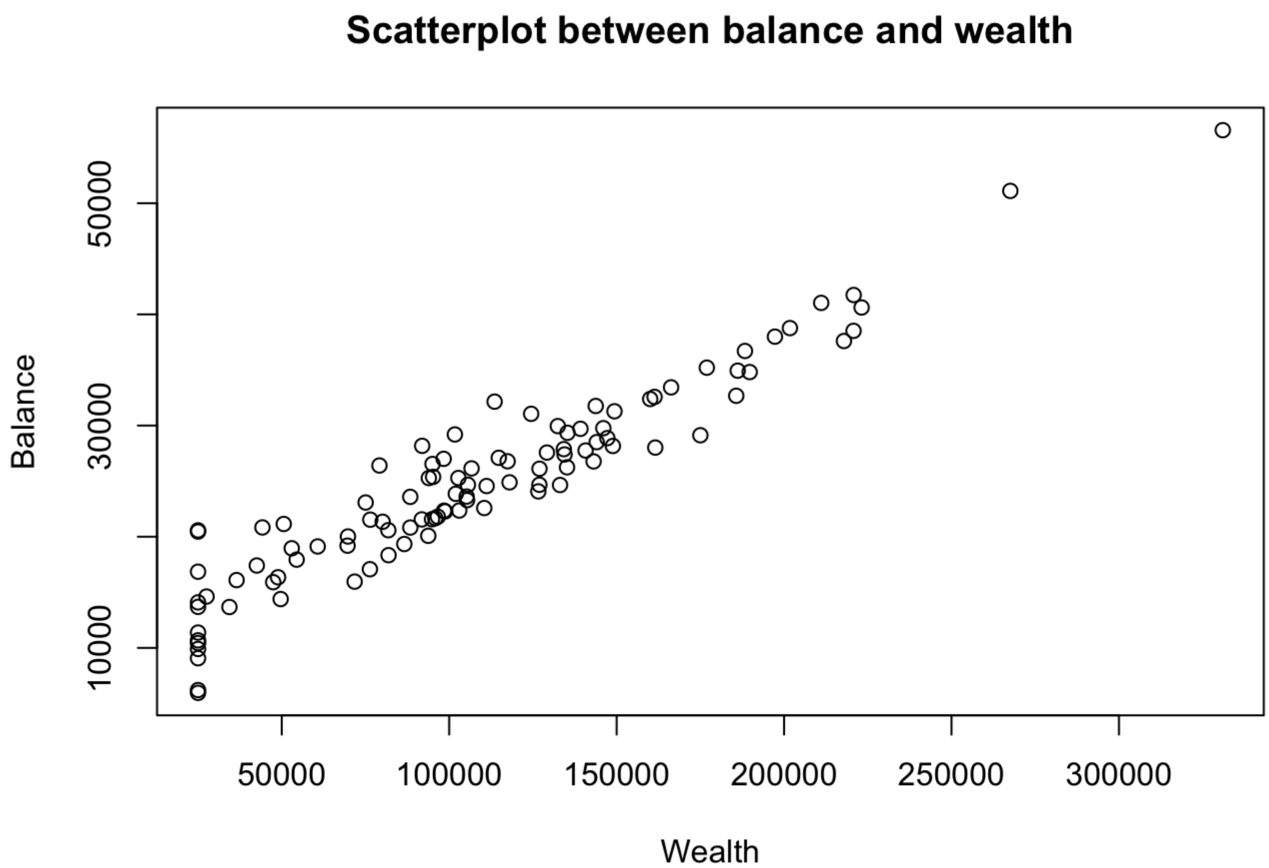
```
# 4 scatterplot between balance and homeval
plot(homeval, balance, main="Scatterplot between balance and homeval", xlab="HomeVal", y
lab="Balance")
```



Answer:

The relationship between balance and homeval does appear to be linear, but not as much as for income.

```
# 5 scatterplot between balance and wealth
plot(wealth, balance, main="Scatterplot between balance and wealth", xlab="Wealth", ylab
="Balance")
```



Answer:

The relationship between balance and wealth can be described as linear. It appears that balance has the most clearly linear associations with wealth and income.

Problem 1 b)

Compute correlation values of bank balance vs the other variables. Interpret the correlation values, and discuss which variables appear to be strongly associated. [1 pt R code, 2 pts for answer = 3 pts]

```
# calculate correlation values between bank balance and the other variables

cor(myd)
```

```
##                 Age Education    Income   HomeVal    Wealth   Balance
## Age       1.0000000 0.1734071 0.4771474 0.3864931 0.4680918 0.5654668
## Education 0.1734071 1.0000000 0.5753940 0.7535211 0.4694130 0.5548807
## Income    0.4771474 0.5753940 1.0000000 0.7953552 0.9466654 0.9516845
## HomeVal   0.3864931 0.7535211 0.7953552 1.0000000 0.6984778 0.7663871
## Wealth    0.4680918 0.4694130 0.9466654 0.6984778 1.0000000 0.9487117
## Balance   0.5654668 0.5548807 0.9516845 0.7663871 0.9487117 1.0000000
```

Answer:

Extremely strong correlations - Balance and Income (0.952), Balance and Wealth (0.949), Wealth and Income (0.947), Strong Correlations - Balance and HomeVal (0.767), Wealth and HomeVal (0.698), HomeVal and Education (0.754), HomeVal and Income (0.795) Other Noteable Correlations - Balance and Age (0.565), Balance and Education (0.555), Income and Education (0.575) ^I used the same line of reasoning as Assignment 1 Solutions for what constitutes a noteable correlation (above 0.55)

Problem 1 c)

Fit a regression model of balance vs the other five variables (model M1). Compute the VIF statistics for each x-variable and analyze whether there is a problem of multicollinearity. [2 pts R code, 1 pt answer = 3 pts]

```
# Fit a regression model (M1) balance v. the other variables
M1 <- lm(balance ~ age+education+income+homeval+wealth, data=myd)

# Compute the VIF statistics
library(car)
```

```
## Loading required package: carData
```

```
vif(M1)
```

```
##       age education    income   homeval    wealth
##  1.342764  2.456706 14.901724  4.382999 10.714276
```

Answer:

Yes, there is multicollinearity with income and wealth because both have VIF >= 10. We have two choices, remove one of the correlated variables or redefine the variables to reduce multicollinearity.

Remove Strategy - Because Wealth and Income are very strongly correlated (0.947), it will not be necessary to include both in our regression model. Income and Balance are more strongly correlated than Wealth and Balance, so it is reasonable to keep income. Redefine Strategy - Standardize the variables

Problem 1 d)

Apply your knowledge of regression analysis to define a better model M2, and answer the following questions:

```
# Analyze the model parameters
summary(M1)
```

```
##
## Call:
## lm(formula = balance ~ age + education + income + homeval + wealth,
##     data = myd)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5376.9 -1110.8   -77.2   872.3  7732.3
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.071e+04  4.261e+03  -2.514 0.013613 *
## age          3.187e+02  6.099e+01   5.225 1.01e-06 ***
## education    6.219e+02  3.190e+02   1.950 0.054135 .
## income       1.463e-01  4.078e-02   3.588 0.000527 ***
## homeval      9.183e-03  1.104e-02   0.832 0.407505
## wealth       7.433e-02  1.119e-02   6.643 1.85e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2056 on 96 degrees of freedom
## Multiple R-squared:  0.9469, Adjusted R-squared:  0.9441
## F-statistic: 342.4 on 5 and 96 DF,  p-value: < 2.2e-16
```

```
vif(M1)
```

```
##       age education    income   homeval    wealth
##  1.342764  2.456706 14.901724  4.382999 10.714276
```

Answer:

Homeval is not significant at the 5% significance level, which suggests it can be removed from the regression model and we can refit. HOWEVER, there is multicollinearity which may be throwing this off. First, I will refit with either wealth or income removed because those are the vif>=10 and highly correlated.

```
# Fit a regression model (M2) balance v. the other variables
M2withwealth <- lm(balance ~ age+education+wealth+homeval, data=myd)
M2withincome <- lm(balance ~ age+education+income+homeval, data=myd)

# Compute the VIF statistics
vif(M2withwealth)
```

```
##       age education    wealth   homeval
##  1.329055  2.415324  2.155681  3.696838
```

```
vif(M2withincome)
```

```
##       age education    income   homeval
##  1.340976  2.404509  2.998184  4.291805
```

```r
# Analyze the model parameters
summary(M2withwealth)
```

```
##
## Call:
## lm(formula = balance ~ age + education + wealth + homeval, data = myd)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7586.5 -1090.2    29.8   914.2  7670.9
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.187e+04  4.501e+03  -2.636  0.00976 **
## age          3.408e+02  6.428e+01   5.301 7.22e-07 ***
## education    7.704e+02  3.351e+02   2.299  0.02363 *
## wealth       1.102e-01  5.317e-03  20.727  < 2e-16 ***
## homeval      2.485e-02  1.074e-02   2.314  0.02277 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2178 on 97 degrees of freedom
## Multiple R-squared:  0.9398, Adjusted R-squared:  0.9373
## F-statistic: 378.5 on 4 and 97 DF,  p-value: < 2.2e-16
```

```r
summary(M2withincome)
```

```
##
## Call:
## lm(formula = balance ~ age + education + income + homeval, data = myd)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7645.1 -1560.3   -50.9  1164.1  8432.9
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -9.809e+03  5.119e+03  -1.916   0.0583 .
## age          3.335e+02  7.325e+01   4.552 1.54e-05 ***
## education    3.130e+02  3.793e+02   0.825   0.4112
## income       3.885e-01  2.199e-02  17.668  < 2e-16 ***
## homeval     -1.394e-03  1.313e-02  -0.106   0.9157
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2471 on 97 degrees of freedom
## Multiple R-squared:  0.9225, Adjusted R-squared:  0.9193
## F-statistic: 288.7 on 4 and 97 DF,  p-value: < 2.2e-16
```

Answer:

Now multicollinearity is not a problem as there are no VIF >= 10. The model with age, education, income, and homeval leads to high p values for education and homeval, which isn't good. The model: "lm(formula = balance ~ age + education + wealth + homeval, data = myd)" has all variables at the 5% significance level and no multicollinearity. Adj-R2 is also higher. So: I will use age, education, wealth, and homeval in the revised model.

```
# Fit a regression model (M2) balance v. the other variables
M2 <- lm(formula = balance ~ age + education + wealth + homeval, data = myd)

# Compute the VIF statistics
vif(M2)
```

```
##       age education    wealth   homeval
##  1.329055  2.415324  2.155681  3.696838
```

```
# Analyze the model parameters
summary(M2)
```

```
##
## Call:
## lm(formula = balance ~ age + education + wealth + homeval, data = myd)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7586.5 -1090.2    29.8   914.2  7670.9
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.187e+04  4.501e+03  -2.636  0.00976 **
## age          3.408e+02  6.428e+01   5.301 7.22e-07 ***
## education    7.704e+02  3.351e+02   2.299  0.02363 *
## wealth       1.102e-01  5.317e-03  20.727  < 2e-16 ***
## homeval      2.485e-02  1.074e-02   2.314  0.02277 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2178 on 97 degrees of freedom
## Multiple R-squared:  0.9398, Adjusted R-squared:  0.9373
## F-statistic: 378.5 on 4 and 97 DF,  p-value: < 2.2e-16
```

Answer:

From the summary function, we see that the F-statistic is 378.5 with p-value of 2.2e-16 < 0.001. Therefore, we can reject the null hypothesis that there is no significant effect of age, education, wealth, and homeval on balance (if all the Beta parameters are 0). So we can accept the alternative hypothesis (at least one of the Beta parameters is not 0) and conclude that there is at least one variable that has a significant effect on balance.

d) 1

Analyze the Coefficient of Determination R2 values and the adjusted adj-R2 values for both models M1 and M2. Which model has the largest adj-R2 value? [1 pt selecting better model M2, 1 pt R code, 3 pts answer = 5 pts]

Answer:

M1 Coefficient of Determination R2: 0.9469 ~ 0.95 M1 Adj-R2: 0.9441 ~0.94

M2 Coefficient of Determination R2: 0.9398 ~0.94 M2 Adj-R2: 0.9373 ~0.94

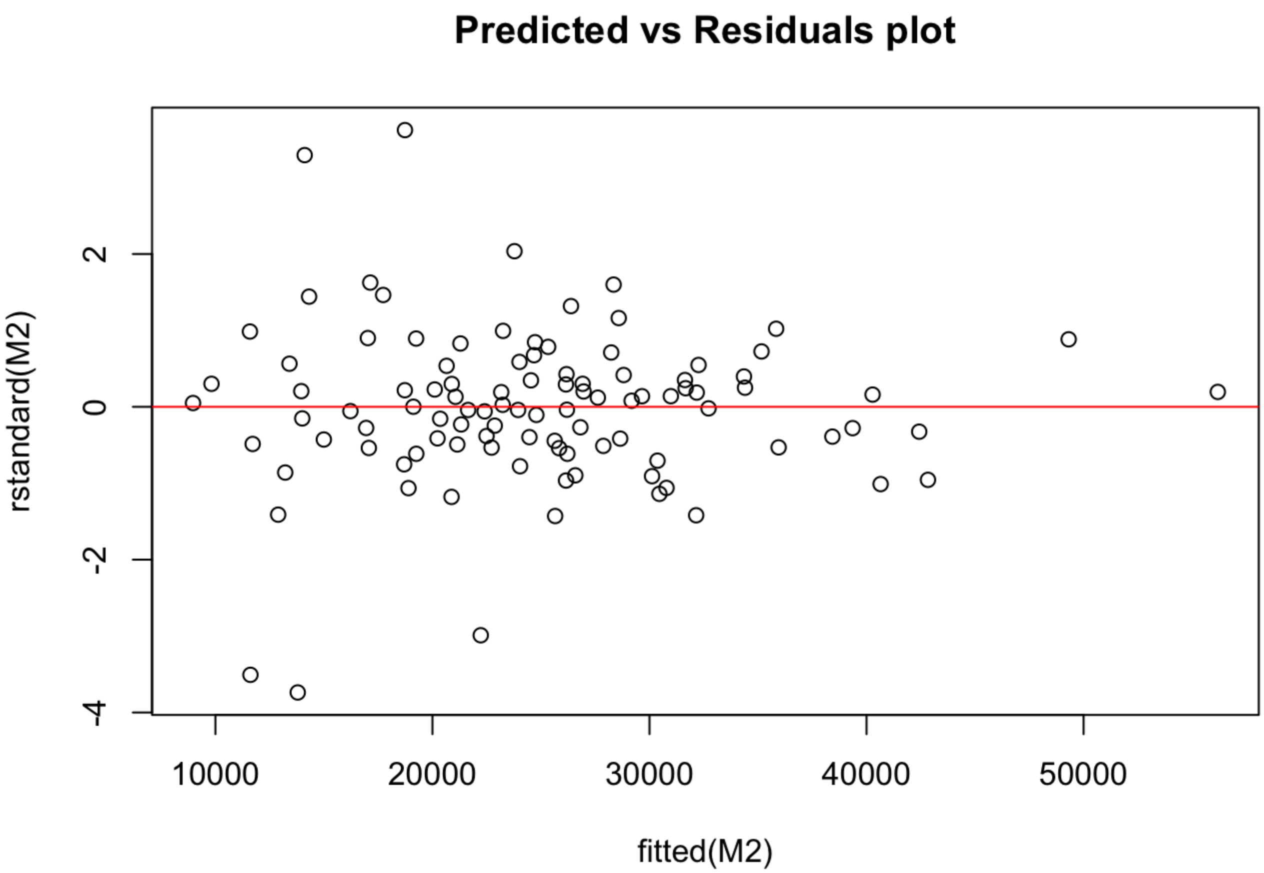
The coefficient of determination R2 represents the amount of variation in Y explained by the regression model. So for this analysis, the M1 R2 is about 0.95 and M2 R2 is about 0.94, which means that about 95% (used age,education,wealth,homeval,income) and 94% (used age,education,wealth,homeval) of the respective variation in bank balance is explained by the relationship with the variables used in each model.

The Adjusted-R2 for M2 is slightly lower than for M1, but because it does not have the multicollinearity problem and it is a simpler model I would argue that it is the better model.

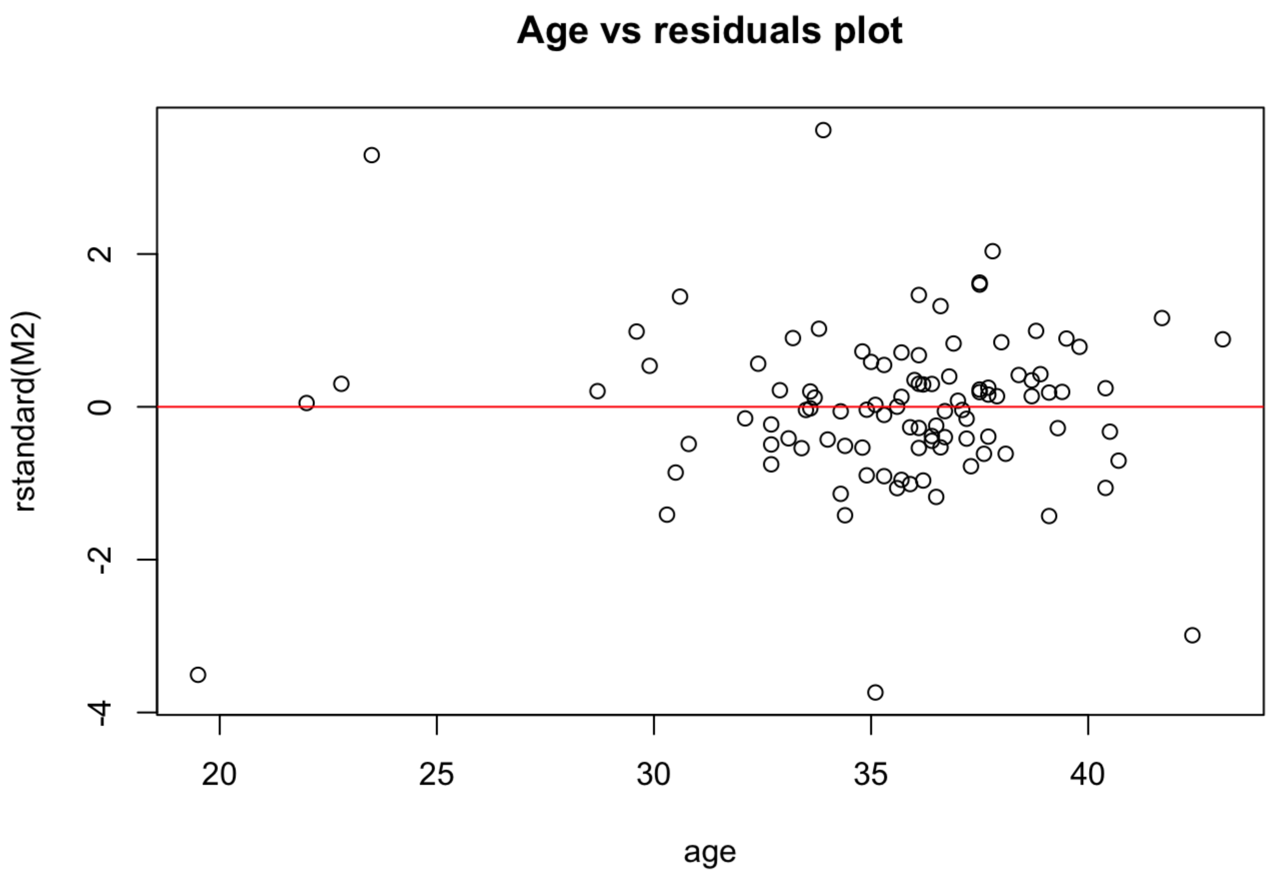
d) 2

Create residual plots (standardized residuals vs predicted; standardized residuals vs x-variables; and normal plot of residuals). Analyze the residual plots to check if the regression model assumptions are met by the data. [1 pt R code, 1 pt plots, 1 pt answer = 3 pts]

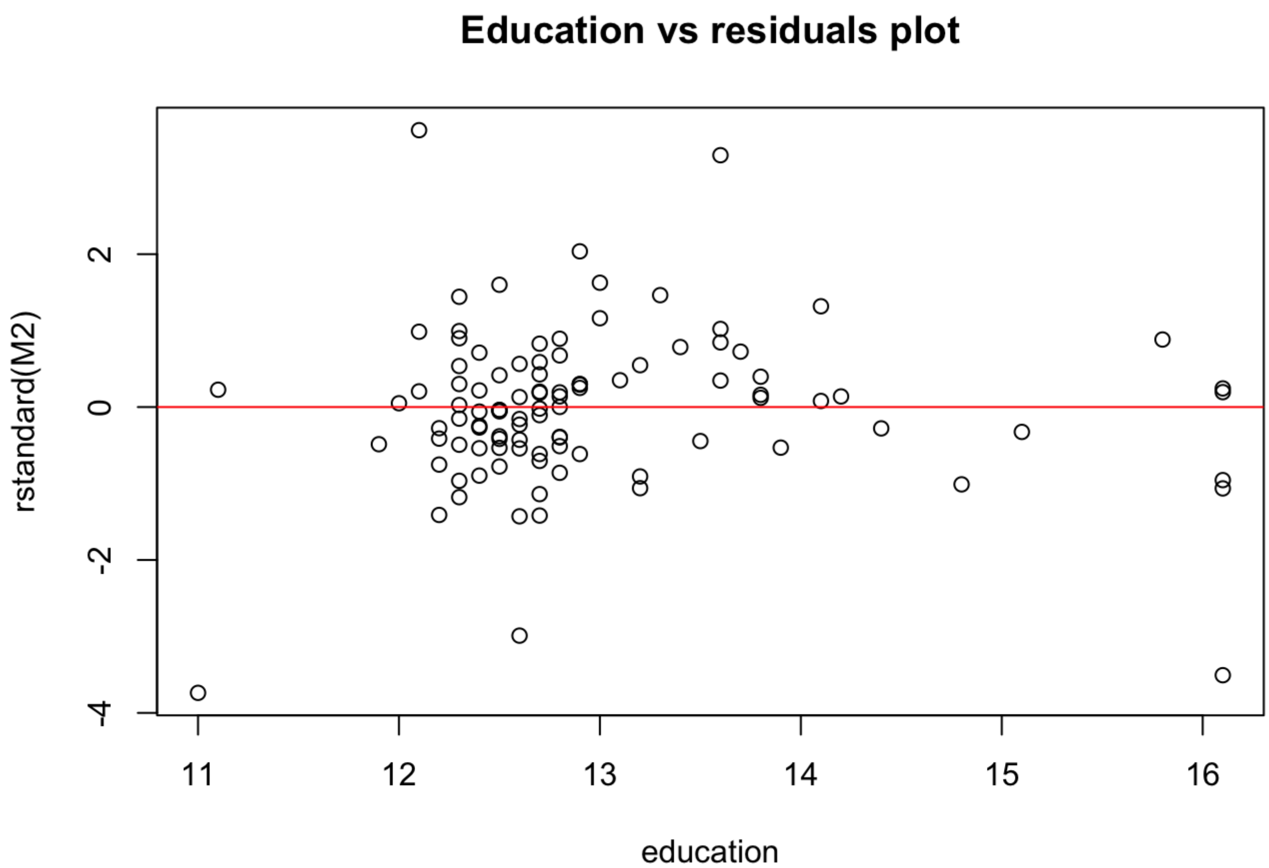
```
#residual plots
#Plot residuals vs predicted values
plot( fitted(M2), rstandard(M2), main="Predicted vs Residuals plot")
abline(a=0, b=0, col='red') #add zero line
```



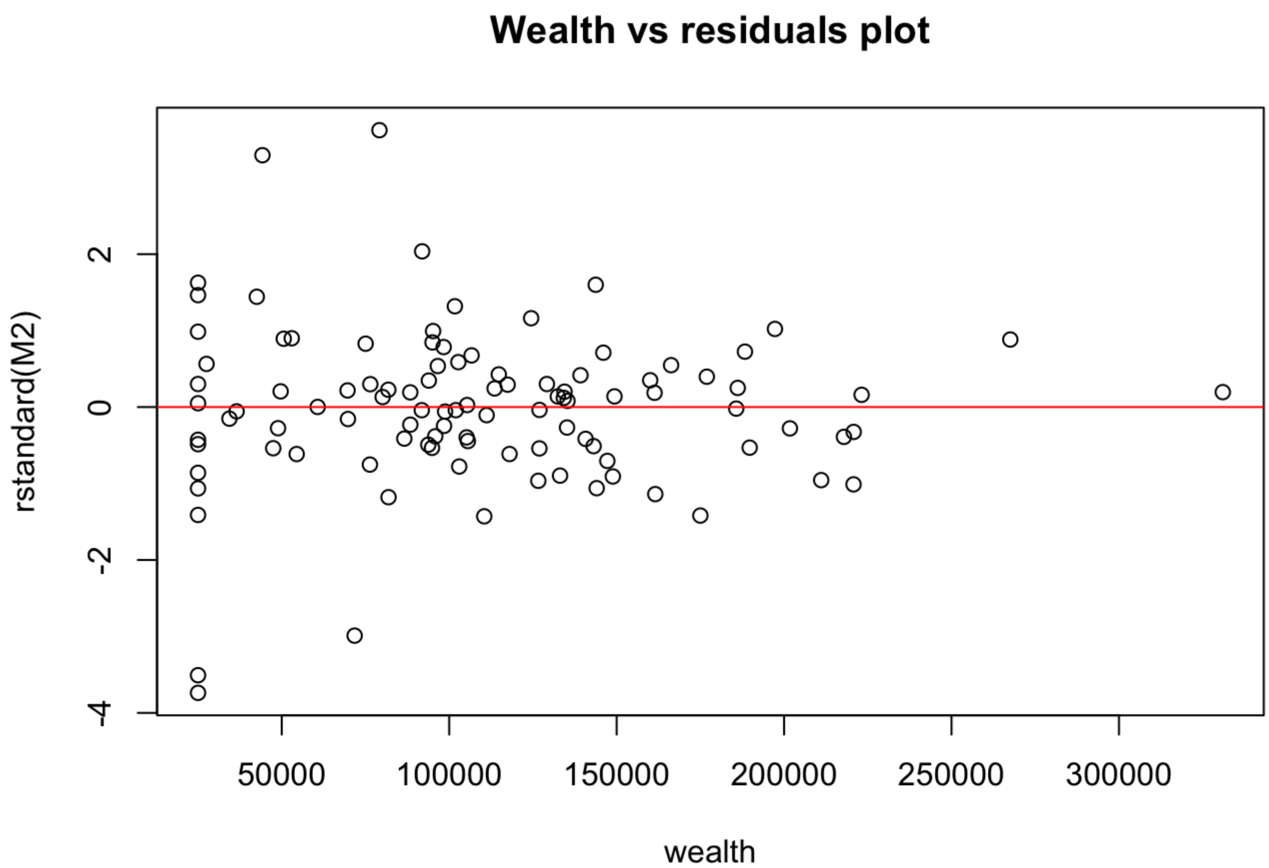

```
#Plot residuals vs each x-variable:
plot(age, rstandard(M2), main="Age vs residuals plot")
abline(a=0, b=0,col='red')
```



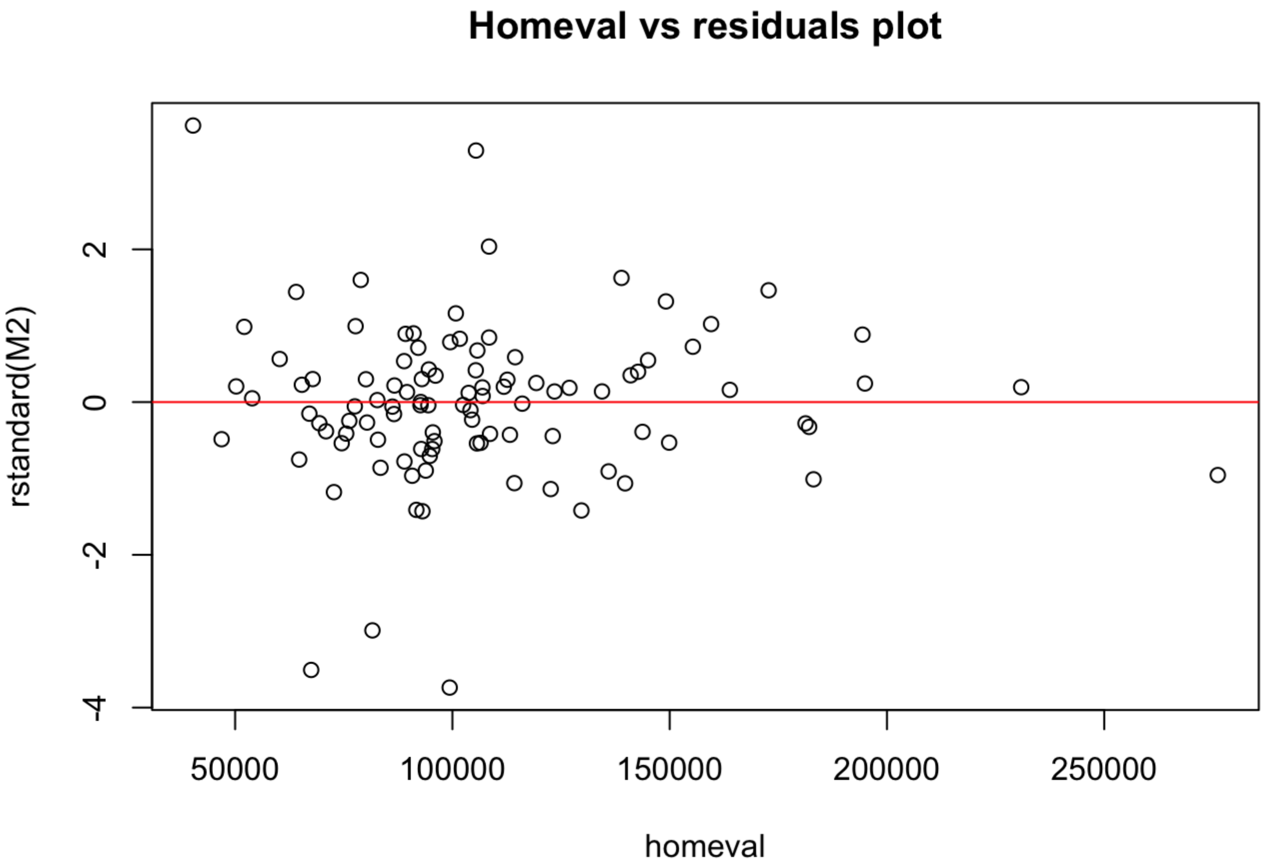

```
plot(education, rstandard(M2), main="Education vs residuals plot")
abline(a=0, b=0,col='red')
```



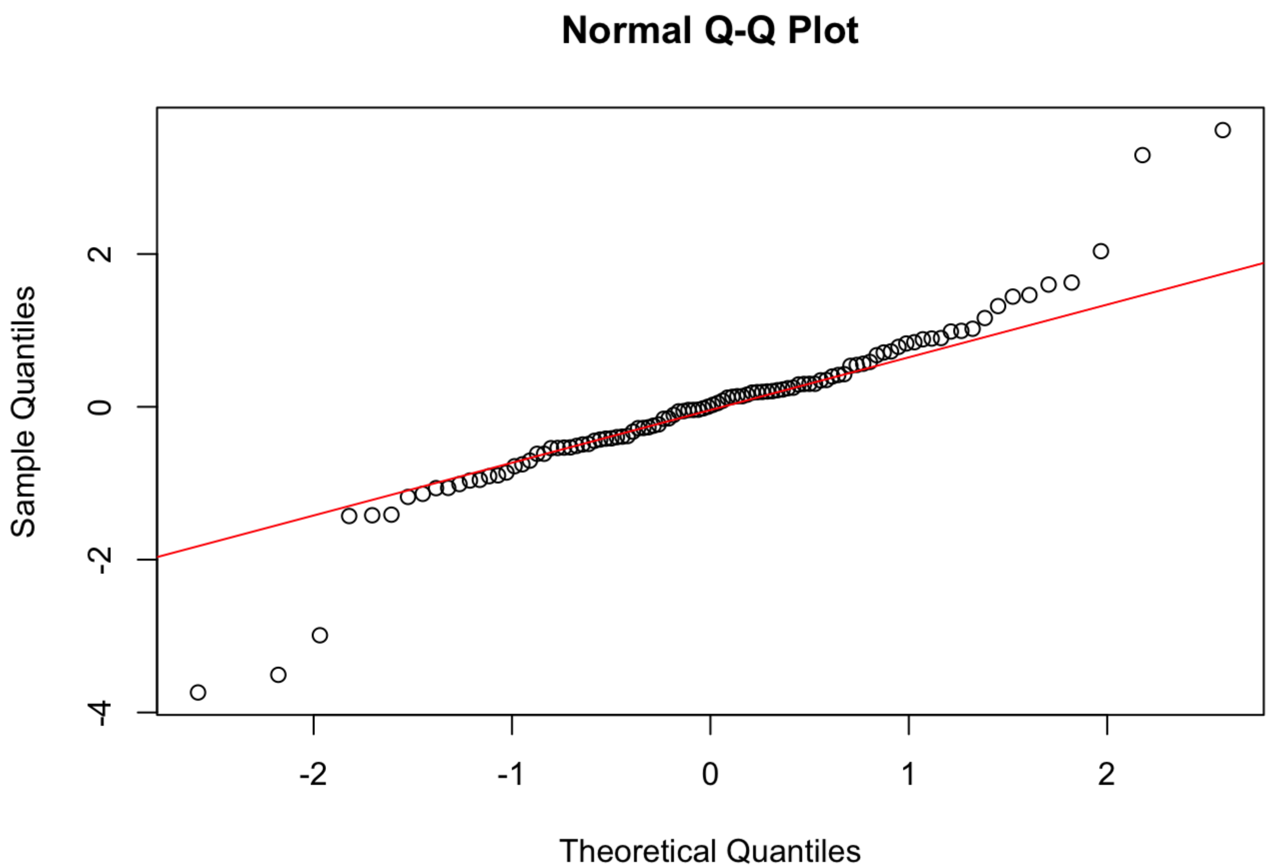

```
plot(wealth, rstandard(M2), main="Wealth vs residuals plot")
abline(a=0, b=0,col='red')
```




```
plot(homeval, rstandard(M2), main="Homeval vs residuals plot")
abline(a=0, b=0,col='red')
```




```
#normal probability plot of residuals
qqnorm(rstandard(M2))
qqline(rstandard(M2), col = 2)
```



Answer: Based on the residual plots, I do believe the model assumptions are met by the data. * Standardized residuals vs predicted: The first plot, predicted v. residuals, does not appear to show a random scatter. That said, the influential points analysis suggests that this may be due to these points, with their removal, it should confirm that the regression model assumptions are met. * Standardized residuals vs x-variables: Age and Education seem find, Wealth and Homevalue appear to be showing a change in variation, which could indicate the model is off or that there are outliers that are throwing it off. * Normal plot of residuals: There are about 5 points that don't fall well along the line. When I check for influential points, these are likely to show up. **Model Assumptions: Linearity:** * Predicted v. residuals – Points not entirely random, linearity not completely satisfied. But with outlier removal I think it would be *Cases v. residuals – Random for age and education so linearity confirmed, not so much for wealth and homeval. **Constant Variance:*** Predicted v. residuals – Doesn't appear to be random, but would be with removal of 5 outliers. **Normality of errors:** *NormalQQplot – Good, points close to line indicating normal distribution of errors. **Outliers?*** Yes, visible on qq plot but also discovered in later analysis.

d) 3

Analyze if there are any outliers and influential points for your model. If so, what are your recommendations? [2 pts answer, 2 pts R code = 4 pts]

```
#run model diagnostics....
# compute influential points statistics
influence.measures(M2)
```

```
## Influence measures of
##   lm(formula = balance ~ age + education + wealth + homeval, data = myd) :
##
##        dfb.1_  dfb.age  dfb.edct  dfb.wlth  dfb.hmvl     dffit cov.r
## 1    1.76e-02  8.99e-02 -4.49e-02 -1.01e-01 -4.60e-02 -0.263443 1.067
## 2    5.76e-03 -7.02e-03 -6.00e-03  2.05e-02  6.78e-03  0.036130 1.106
## 3   -6.74e-03 -7.41e-03  8.97e-03  2.92e-02 -1.57e-04  0.061853 1.070
## 4    5.27e-02 -4.04e-02 -4.81e-02  2.64e-02  5.29e-02  0.098976 1.071
## 5   -6.09e-02  4.83e-02  5.41e-02 -2.55e-02 -6.61e-02 -0.133230 1.031
## 6    5.05e-02 -7.33e-02 -3.20e-02  6.26e-02  5.08e-02  0.147265 1.067
## 7    8.49e-03 -5.38e-03 -3.52e-04 -6.30e-03 -2.34e-02 -0.061957 1.101
## 8    2.50e-03  1.99e-02 -7.96e-03 -4.78e-02 -5.12e-03 -0.093290 1.070
## 9   -7.30e-02  1.50e-01  8.09e-02  1.61e-01 -4.19e-01 -0.557496 1.347
## 10   4.69e-02 -1.95e-02 -3.86e-02 -2.13e-02  5.79e-03 -0.086701 1.123
## 11  -2.14e-02  1.02e-02  2.07e-02  1.04e-03 -1.24e-02  0.026593 1.091
## 12  -2.24e-01  1.04e-01  1.91e-01  1.25e-01 -9.68e-02  0.339882 1.162
## 13   6.83e-03 -1.19e-04 -7.89e-03  3.48e-02 -9.52e-03  0.046533 1.086
## 14   2.98e-02 -1.66e-02 -3.05e-02  1.26e-02  3.19e-02  0.058246 1.076
## 15  -6.87e-02  3.81e-02  5.50e-02 -5.80e-02  1.97e-02  0.105940 1.251
## 16   1.19e-01 -1.49e-01 -8.04e-02  1.09e-01  9.92e-02  0.249541 1.057
## 17   1.68e-02 -1.30e-02 -1.29e-02  2.31e-02 -1.11e-02 -0.055286 1.059
## 18  -4.59e-02  1.88e-02  4.96e-02 -6.11e-02 -2.17e-02 -0.098548 1.112
## 19   6.14e-03  2.61e-03 -8.01e-03 -3.33e-04  6.72e-03  0.030735 1.060
## 20   8.73e-03  8.68e-03 -1.60e-02  8.25e-03  8.35e-03  0.031757 1.082
## 21  -2.14e-02 -1.47e-02  2.50e-02  4.92e-02 -2.70e-03  0.089409 1.274
## 22   1.02e-02  1.15e-02 -1.47e-02 -6.97e-03  8.98e-03  0.070163 1.040
## 23  -4.11e-03  1.15e-03  3.69e-03 -1.09e-03 -1.55e-03 -0.011210 1.065
## 24  -1.83e-03  1.35e-02 -4.18e-03 -1.34e-02  6.60e-03  0.026009 1.071
## 25  -1.24e-02  2.56e-02 -4.73e-03 -6.07e-02  3.69e-02 -0.081126 1.065
## 26  -9.75e-03 -5.54e-03  1.60e-02  1.13e-02 -1.50e-02  0.023211 1.092
## 27   8.91e-02 -1.30e-01 -4.14e-02 -2.98e-02  6.06e-02 -0.183541 1.023
## 28  -2.11e-02  4.37e-02  6.75e-03  5.60e-03 -2.43e-02  0.065363 1.068
## 29  -1.36e-01  1.07e-01  1.09e-01 -1.11e-01 -5.64e-02 -0.210790 1.018
## 30   8.19e-03  5.41e-03 -1.39e-02  2.57e-05  1.24e-02  0.023663 1.083
## 31  -5.04e-02  4.45e-02  3.38e-02 -3.27e-02 -1.26e-02 -0.078752 1.060
## 32   4.19e-02 -1.36e-02 -4.28e-02 -1.98e-02  4.58e-02  0.076305 1.052
## 33   2.45e-02 -4.78e-02 -1.13e-02 -1.98e-02  4.00e-02 -0.087979 1.054
## 34  -2.69e-03  1.40e-03  2.24e-03 -2.05e-03 -3.75e-04 -0.004916 1.071
## 35  -8.99e-03  4.68e-03  1.22e-02  2.35e-02 -2.67e-02  0.042204 1.069
## 36  -2.18e-02  1.73e-02  1.58e-02  5.90e-03 -1.61e-02 -0.033061 1.072
## 37  -5.54e-05 -2.20e-03  8.00e-04  1.92e-05  7.87e-04 -0.005047 1.069
## 38   4.10e-01 -1.08e+00  1.12e-01 -1.78e-01  2.70e-01  1.298258 0.663
## 39  -1.04e-01  8.51e-02  8.54e-02 -4.56e-02 -4.56e-02  0.143658 1.044
## 40   1.90e-02 -1.70e-02 -1.32e-02  1.26e-02  6.64e-03  0.029948 1.075
## 41  -1.42e-01  2.14e-01  6.71e-02  7.96e-04 -1.00e-01  0.254449 1.029
## 42  -6.90e-02  4.01e-02  4.55e-02 -1.65e-01  1.27e-01  0.264508 1.001
## 43  -5.65e-02  3.29e-02  4.06e-02 -8.14e-02  2.67e-02 -0.133062 1.033
## 44   9.58e-03 -1.79e-02 -6.20e-03 -2.69e-03  1.54e-02 -0.045548 1.059
## 45   9.02e-03  2.31e-02 -2.09e-02  2.00e-02 -4.06e-03  0.062015 1.067
## 46  -4.18e-02  1.15e-02  4.22e-02  1.78e-02 -3.76e-02 -0.070068 1.056
## 47  -2.40e-03  2.55e-03  1.47e-03 -3.65e-03  2.06e-04 -0.004728 1.111
## 48  -1.37e-02  4.56e-03  1.49e-02  5.09e-03 -1.29e-02  0.017804 1.106
## 49  -2.01e-01  1.53e-01  1.68e-01 -1.60e-01 -9.66e-02 -0.303137 0.991
```

```
## 50  -2.16e-02 -6.51e-03  2.79e-02 -2.23e-02 -4.61e-03 -0.058846 1.065
## 51  -1.29e-04  4.33e-03 -5.50e-04 -4.36e-03 -1.37e-03  0.015247 1.067
## 52  -4.88e-02  1.62e-01 -1.96e-02 -1.39e-01  5.27e-02  0.282503 0.862
## 53  -3.00e-05  6.40e-05  1.71e-05 -1.19e-04  1.22e-05  0.000193 1.074
## 54   3.33e-02 -1.45e-02 -2.37e-02  8.57e-02 -4.20e-02  0.119518 1.055
## 55  -2.66e-03  1.39e-03  1.63e-03 -1.81e-03  1.28e-03 -0.007157 1.068
## 56  -1.20e-01  1.22e-01  8.68e-02 -3.19e-02 -7.20e-02  0.161333 1.063
## 57  -4.25e-02 -6.59e-03  4.32e-02 -6.81e-02  3.25e-02 -0.138639 1.024
## 58   9.52e-04 -9.86e-05 -6.88e-04  1.25e-03 -1.07e-03  0.003432 1.071
## 59   4.20e-01 -1.37e-01 -4.30e-01  3.10e-01  7.49e-02 -0.571145 1.279
## 60   5.26e-02 -9.47e-02 -2.09e-02 -5.49e-02  7.55e-02 -0.147763 1.072
## 61  -2.73e-03  1.93e-03  1.85e-03 -1.66e-05 -8.86e-04 -0.004949 1.069
## 62  -5.86e-02  8.07e-02  5.86e-02  2.18e-01 -2.32e-01  0.332617 0.961
## 63   3.95e-02 -6.81e-02 -1.71e-02  7.68e-02 -5.20e-03 -0.112014 1.067
## 64   6.38e-02 -1.59e-01 -8.06e-03 -4.97e-03  7.10e-02 -0.227958 0.971
## 65  -2.52e-02  1.91e-02  1.26e-02 -1.46e-02  1.30e-02 -0.058038 1.065
## 66  -2.44e-02  2.18e-02  2.31e-02 -6.43e-03 -2.65e-02  0.046811 1.074
## 67   9.51e-03 -4.88e-02  4.66e-03 -7.54e-03  3.12e-02 -0.102098 1.038
## 68  -5.93e-02  4.65e-02  5.10e-02 -1.09e-02 -4.11e-02  0.073552 1.094
## 69  -2.14e-03  5.32e-02 -2.06e-02 -7.12e-02  3.25e-02  0.117261 1.037
## 70  -3.75e-02  3.30e-02  2.00e-02 -1.92e-02  6.27e-03 -0.069213 1.061
## 71   1.56e-02 -4.12e-02 -6.69e-03  4.39e-02  1.36e-02 -0.088162 1.066
## 72  -9.20e-02  1.57e-01  3.81e-02 -1.23e-01 -1.54e-02  0.204144 1.063
## 73   2.42e-02 -7.50e-02 -1.26e-02 -4.04e-03  8.16e-02 -0.176748 1.002
## 74   1.41e-02 -9.79e-03 -8.89e-03 -6.58e-03  5.32e-03  0.028705 1.069
## 75   7.71e-02 -8.54e-02 -4.03e-02  2.95e-02  1.56e-02  0.107800 1.080
## 76  -4.69e-03  4.44e-03 -2.61e-05  9.40e-03  2.08e-03 -0.024891 1.081
## 77   1.93e-01  4.37e-02 -3.04e-01 -6.16e-01  6.49e-01  0.748762 1.185
## 78  -2.17e-03 -1.04e-03 -1.63e-03 -3.26e-02  2.82e-02 -0.047343 1.082
## 79  -2.71e-02  3.11e-02 -1.90e-03 -3.09e-02  5.31e-02 -0.117946 1.048
## 80   5.85e-03 -1.37e-02 -1.82e-03  1.00e-02  3.21e-03 -0.023214 1.075
## 81  -4.40e-02  1.18e-01  7.34e-03  4.16e-03 -7.33e-02  0.177461 1.033
## 82   6.94e-02  1.58e-01 -1.90e-01 -5.28e-01  4.36e-01  0.595652 1.038
## 83   1.86e-02 -2.20e-02 -1.90e-02 -1.73e-02  4.44e-02 -0.063026 1.074
## 84   4.55e-01 -7.91e-01 -1.76e-01  2.92e-01  2.44e-01 -0.883227 0.702
## 85   1.36e+00  1.50e+00 -2.62e+00 -2.99e-01  1.67e+00 -3.385618 0.974
## 86  -2.96e-02  6.27e-02 -1.09e-02  9.36e-02 -3.05e-02 -0.170281 1.054
## 87   7.46e-02 -2.82e-02 -6.57e-02 -7.78e-02  6.88e-02  0.144683 1.036
## 88   2.65e-03 -5.75e-03 -1.06e-03  6.78e-03  5.92e-04 -0.010608 1.091
## 89  -2.04e-02 -9.20e-04  3.99e-02 -3.46e-02 -4.12e-02  0.106128 1.073
## 90   5.17e-03 -1.26e-02 -4.15e-03 -8.93e-03  2.03e-02 -0.036159 1.073
## 91  -1.54e-01  4.73e-02  3.02e-01  3.90e-01 -7.56e-01  0.926675 0.539
## 92   5.80e-02 -8.86e-02  6.10e-03 -2.81e-02 -4.17e-02  0.203971 1.044
## 93   2.44e-02  1.58e-02 -3.35e-02  2.31e-04  3.76e-03  0.053050 1.108
## 94  -3.67e-02  5.96e-05  4.47e-02  9.38e-02 -8.10e-02 -0.112581 1.116
## 95   1.52e-02 -2.72e-02  2.49e-03  9.31e-03 -1.60e-02  0.046344 1.105
## 96   4.33e-03 -2.07e-02  2.41e-04  1.85e-02  8.72e-03 -0.045729 1.078
## 97   7.32e-02 -1.12e-01  6.45e-03 -3.15e-02 -4.52e-02  0.256010 0.974
## 98   6.11e-02 -9.50e-02 -1.59e-02 -2.66e-03  1.65e-02  0.109642 1.188
## 99  -2.07e-01  1.46e-01  1.58e-01  1.93e-01 -2.04e-01 -0.341586 1.004
## 100  1.10e-02 -1.68e-02 -2.81e-03  1.60e-03  1.00e-03  0.019336 1.214
## 101 -2.45e-02  2.62e-02  6.07e-04  1.33e-02  2.59e-02 -0.096694 1.082
## 102 -1.05e+00 -1.59e-02  1.27e+00  8.79e-01 -1.20e+00 -1.564027 0.557
##       cook.d    hat inf
```

```
## 1   1.39e-02 0.0636
## 2   2.64e-04 0.0489
## 3   7.72e-04 0.0239
## 4   1.97e-03 0.0318
## 5   3.56e-03 0.0211
## 6   4.36e-03 0.0398
## 7   7.75e-04 0.0474
## 8   1.75e-03 0.0302
## 9   6.22e-02 0.2541   *
## 10  1.52e-03 0.0670
## 11  1.43e-04 0.0357
## 12  2.32e-02 0.1292   *
## 13  4.37e-04 0.0337
## 14  6.85e-04 0.0271
## 15  2.27e-03 0.1604   *
## 16  1.24e-02 0.0563
## 17  6.16e-04 0.0153
## 18  1.96e-03 0.0606
## 19  1.91e-04 0.0110
## 20  2.04e-04 0.0285
## 21  1.61e-03 0.1749   *
## 22  9.90e-04 0.0107
## 23  2.54e-05 0.0112
## 24  1.37e-04 0.0182
## 25  1.33e-03 0.0247
## 26  1.09e-04 0.0363
## 27  6.73e-03 0.0290
## 28  8.62e-04 0.0230
## 29  8.86e-03 0.0330
## 30  1.13e-04 0.0282
## 31  1.25e-03 0.0209
## 32  1.17e-03 0.0167
## 33  1.56e-03 0.0202
## 34  4.88e-06 0.0170
## 35  3.60e-04 0.0194
## 36  2.21e-04 0.0204
## 37  5.15e-06 0.0152
## 38  3.02e-01 0.1223   *
## 39  4.14e-03 0.0281
## 40  1.81e-04 0.0219
## 41  1.29e-02 0.0456
## 42  1.39e-02 0.0384
## 43  3.55e-03 0.0216
## 44  4.19e-04 0.0131
## 45  7.76e-04 0.0219
## 46  9.89e-04 0.0171
## 47  4.52e-06 0.0523
## 48  6.41e-05 0.0478
## 49  1.82e-02 0.0432
## 50  6.99e-04 0.0198
## 51  4.70e-05 0.0135
## 52  1.54e-02 0.0183
## 53  7.53e-09 0.0192
## 54  2.87e-03 0.0276
```

```
## 55  1.03e-05 0.0145
## 56  5.23e-03 0.0407
## 57  3.85e-03 0.0203
## 58  2.38e-06 0.0170
## 59  6.52e-02 0.2234   *
## 60  4.39e-03 0.0424
## 61  4.95e-06 0.0145
## 62  2.18e-02 0.0408
## 63  2.53e-03 0.0323
## 64  1.03e-02 0.0246
## 65  6.80e-04 0.0196
## 66  4.42e-04 0.0241
## 67  2.09e-03 0.0170
## 68  1.09e-03 0.0435
## 69  2.76e-03 0.0197
## 70  9.66e-04 0.0195
## 71  1.57e-03 0.0263
## 72  8.35e-03 0.0497
## 73  6.22e-03 0.0219
## 74  1.66e-04 0.0173
## 75  2.34e-03 0.0391
## 76  1.25e-04 0.0265
## 77  1.11e-01 0.2054   *
## 78  4.53e-04 0.0306
## 79  2.79e-03 0.0241
## 80  1.09e-04 0.0218
## 81  6.30e-03 0.0309
## 82  6.97e-02 0.1165
## 83  8.02e-04 0.0267
## 84  1.43e-01 0.0741   *
## 85  2.02e+00 0.4512   *
## 86  5.81e-03 0.0378
## 87  4.19e-03 0.0252
## 88  2.27e-05 0.0347
## 89  2.27e-03 0.0344
## 90  2.64e-04 0.0213
## 91  1.50e-01 0.0541   *
## 92  8.32e-03 0.0411
## 93  5.68e-04 0.0523
## 94  2.56e-03 0.0653
## 95  4.34e-04 0.0490
## 96  4.22e-04 0.0269
## 97  1.30e-02 0.0302
## 98  2.43e-03 0.1180   *
## 99  2.31e-02 0.0548
## 100 7.56e-05 0.1324   *
## 101 1.88e-03 0.0384
## 102 4.23e-01 0.1315   *
```

```
# print out only observations that may be influential
summary(influence.measures(M2))
```

```
## Potentially influential observations of
##   lm(formula = balance ~ age + education + wealth + homeval, data = myd) :
##
##     dfb.1_  dfb.age dfb.edct dfb.wlth dfb.hmvl dffit   cov.r   cook.d
## 9   -0.07    0.15    0.08     0.16    -0.42    -0.56    1.35_*  0.06
## 12  -0.22    0.10    0.19     0.13    -0.10     0.34    1.16_*  0.02
## 15  -0.07    0.04    0.05    -0.06     0.02     0.11    1.25_*  0.00
## 21  -0.02   -0.01    0.03     0.05     0.00     0.09    1.27_*  0.00
## 38   0.41   -1.08_*  0.11    -0.18     0.27     1.30_*  0.66_*  0.30
## 59   0.42   -0.14   -0.43     0.31     0.07    -0.57    1.28_*  0.07
## 77   0.19    0.04   -0.30    -0.62     0.65     0.75_*  1.19_*  0.11
## 84   0.46   -0.79   -0.18     0.29     0.24    -0.88_*  0.70_*  0.14
## 85   1.36_*  1.50_* -2.62_*  -0.30     1.67_*  -3.39_*  0.97    2.02_*
## 91  -0.15    0.05    0.30     0.39    -0.76     0.93_*  0.54_*  0.15
## 98   0.06   -0.10   -0.02     0.00     0.02     0.11    1.19_*  0.00
## 100  0.01   -0.02    0.00     0.00     0.00     0.02    1.21_*  0.00
## 102 -1.05_* -0.02    1.27_*   0.88    -1.20_*  -1.56_*  0.56_*  0.42
##     hat
## 9    0.25_*
## 12   0.13
## 15   0.16_*
## 21   0.17_*
## 38   0.12
## 59   0.22_*
## 77   0.21_*
## 84   0.07
## 85   0.45_*
## 91   0.05
## 98   0.12
## 100  0.13
## 102  0.13
```

Answer: There appear to be a number of potentially influencial points. For example in row 85, the data point has a high Cook distance >1, so it is flagged.

```
# compute for all
dfbeta(M2)
```

```
##       (Intercept)          age     education        wealth       homeval
## 1      79.2901950   5.777291133 -1.503169e+01 -5.364502e-04 -4.938725e-04
## 2      26.0662950  -0.453380602 -2.020802e+00  1.093988e-04  7.320420e-05
## 3     -30.4611939  -0.478578269  3.019222e+00  1.558135e-04 -1.689731e-06
## 4     238.2311182  -2.608046603 -1.618344e+01  1.407781e-04  5.698460e-04
## 5    -274.4055805   3.105248946  1.813490e+01 -1.357210e-04 -7.105826e-04
## 6     228.1024174  -4.722985946 -1.076252e+01  3.338684e-04  5.471877e-04
## 7      38.4117293  -0.347163170 -1.183752e-01 -3.365411e-05 -2.527046e-04
## 8      11.3136370   1.287135852 -2.677288e+00 -2.549625e-04 -5.516553e-05
## 9    -328.6059893   9.620099951  2.712970e+01  8.539891e-04 -4.505697e-03
## 10    211.9572157  -1.262207092 -1.298889e+01 -1.140180e-04  6.246763e-05
## 11    -96.9995023   0.657353986  6.961496e+00  5.559894e-06 -1.336092e-04
## 12  -1009.4500478   6.669315336  6.405476e+01  6.671473e-04 -1.040361e-03
## 13     30.9000425  -0.007692799 -2.656472e+00  1.859670e-04 -1.027354e-04
## 14    134.8053544  -1.070585397 -1.026633e+01  6.704704e-05  3.436976e-04
## 15   -310.7190673   2.459654733  1.850390e+01 -3.097446e-04  2.126756e-04
## 16    533.9048827  -9.557985276 -2.692715e+01  5.792304e-04  1.065588e-03
## 17     75.7798375  -0.837603136 -4.333904e+00  1.233158e-04 -1.200832e-04
## 18   -207.5891297   1.214107610  1.667562e+01 -3.261585e-04 -2.338622e-04
## 19     27.7830429   0.168516405 -2.697662e+00 -1.777481e-06  7.255914e-05
## 20     39.4895371   0.560827848 -5.373811e+00  4.409634e-05  9.013281e-05
## 21    -96.7975263  -0.948606263  8.432324e+00  2.627888e-04 -2.915917e-05
## 22     45.8577492   0.744411207 -4.931338e+00 -3.718839e-05  9.673403e-05
## 23    -18.6100947   0.074000054  1.243293e+00 -5.847678e-06 -1.673231e-05
## 24     -8.3007296   0.872894936 -1.408209e+00 -7.179822e-05  7.119198e-05
## 25    -56.0263697   1.649312677 -1.591284e+00 -3.239108e-04  3.975590e-04
## 26    -44.1031386  -0.357848149  5.402556e+00  6.020275e-05 -1.619300e-04
## 27    400.9734279  -8.352434090 -1.387039e+01 -1.582314e-04  6.503007e-04
## 28    -95.5712211   2.821046304  2.272558e+00  2.991786e-05 -2.618290e-04
## 29   -611.2590914   6.837140537  3.641398e+01 -5.875208e-04 -6.043903e-04
## 30     37.0409479   0.349818949 -4.675337e+00  1.373050e-07  1.335177e-04
## 31   -227.7044573   2.872975439  1.137904e+01 -1.743480e-04 -1.361816e-04
## 32    189.4001982  -0.878445452 -1.438558e+01 -1.056549e-04  4.931504e-04
## 33    110.4741615  -3.080712711 -3.782320e+00 -1.055270e-04  4.315076e-04
## 34    -12.1896040   0.090676242  7.543284e-01 -1.096518e-05 -4.048046e-06
## 35    -40.6444366   0.302079956  4.120569e+00  1.256441e-04 -2.879680e-04
## 36    -98.4692451   1.117078914  5.312241e+00  3.153278e-05 -1.732455e-04
## 37     -0.2508028  -0.142290414  2.695046e-01  1.027379e-07  8.493142e-06
## 38   1747.4111577 -65.710665368  3.556377e+01 -8.990110e-04  2.747152e-03
## 39   -469.0955835   5.476059147  2.866243e+01 -2.428390e-04 -4.903096e-04
## 40     86.1347501  -1.098405023 -4.454435e+00  6.718288e-05  7.165507e-05
## 41   -637.8055031  13.727179910  2.242580e+01  4.225744e-06 -1.077065e-03
## 42   -309.5590347   2.568316504  1.518681e+01 -8.724241e-04  1.358546e-03
## 43   -254.6321169   2.117531135  1.361215e+01 -4.333695e-04  2.875663e-04
## 44     43.3166259  -1.154248386 -2.085493e+00 -1.437712e-05  1.658590e-04
## 45     40.7894346   1.488117702 -7.048139e+00  1.066638e-04 -4.378038e-05
## 46   -189.0116264   0.743609894  1.417878e+01  9.486381e-05 -4.049730e-04
## 47    -10.8401030   0.164860873  4.963140e-01 -1.952844e-05  2.218589e-06
## 48    -62.1356045   0.294753842  5.029202e+00  2.719101e-05 -1.387678e-04
## 49   -899.8921999   9.752646497  5.589351e+01 -8.464622e-04 -1.031709e-03
## 50    -97.6451676  -0.420552151  9.386098e+00 -1.189164e-04 -4.974233e-05
## 51     -0.5847036   0.279790320 -1.851617e-01 -2.329188e-05 -1.476755e-05
## 52   -216.1073709  10.254655088 -6.471424e+00 -7.243947e-04  5.564100e-04
```

```
## 53      -0.1358172   0.004132978  5.762112e-03 -6.386346e-07  1.315402e-07
## 54     150.3553589  -0.933847870 -7.956753e+00  4.567350e-04 -4.524457e-04
## 55     -12.0245500   0.089651455  5.501152e-01 -9.655858e-06  1.386437e-05
## 56    -539.2231262   7.882220343  2.913230e+01 -1.698806e-04 -7.751490e-04
## 57    -191.4252750  -0.423617869  1.446992e+01 -3.624776e-04  3.490479e-04
## 58       4.3084588  -0.006374170 -2.318522e-01  6.667762e-06 -1.155320e-05
## 59    1888.0356801  -8.785080000 -1.438775e+02  1.648465e-03  8.041219e-04
## 60     237.2676796  -6.104442283 -7.023304e+00 -2.929333e-04  8.133414e-04
## 61     -12.3549694   0.124625088  6.228669e-01 -8.893501e-08 -9.563184e-06
## 62    -261.4535226   5.144383801  1.946160e+01  1.152066e-03 -2.474001e-03
## 63     178.6020623  -4.394430741 -5.764622e+00  4.097571e-04 -5.598019e-05
## 64     285.7172408 -10.152571293 -2.686721e+00 -2.627834e-05  7.587172e-04
## 65    -113.9944756   1.229978377  4.234859e+00 -7.821918e-05  1.407377e-04
## 66    -110.2665341   1.409927188  7.778810e+00 -3.435113e-05 -2.854821e-04
## 67      42.9166163  -3.143631336  1.566070e+00 -4.015312e-05  3.353639e-04
## 68    -268.1030499   3.002973507  1.716933e+01 -5.820344e-05 -4.432611e-04
## 69      -9.6413465   3.425462543 -6.925503e+00 -3.790802e-04  3.493170e-04
## 70    -169.6044639   2.131997140  6.730032e+00 -1.027405e-04  6.760434e-05
## 71      70.6271589  -2.660617824 -2.251158e+00  2.343817e-04  1.471289e-04
## 72    -414.5617943  10.082273029  1.276913e+01 -6.547544e-04 -1.657001e-04
## 73     108.7682869  -4.812094898 -4.210909e+00 -2.145651e-05  8.740742e-04
## 74      63.6584736  -0.632502613 -2.992804e+00 -3.518009e-05  5.744538e-05
## 75     348.1484858  -5.512047510 -1.356774e+01  1.576507e-04  1.676798e-04
## 76     -21.1983488   0.286576581 -8.775667e-03  5.024942e-05  2.241883e-05
## 77     865.4620613   2.791526358 -1.012450e+02 -3.253318e-03  6.924336e-03
## 78      -9.8170772  -0.067261258 -5.472399e-01 -1.741945e-04  3.041189e-04
## 79    -122.0777816   2.001839270 -6.385703e-01 -1.648571e-04  5.719785e-04
## 80      26.4812268  -0.886433217 -6.134991e-01  5.351945e-05  3.460523e-05
## 81    -197.8855600   7.593788952  2.459560e+00  2.211745e-05 -7.872507e-04
## 82     309.6755026  10.054495367 -6.326491e+01 -2.782765e-03  4.638413e-03
## 83      83.9823932  -1.418574384 -6.397275e+00 -9.259701e-05  4.790881e-04
## 84    1962.2792355 -48.674254719 -5.647343e+01  1.485516e-03  2.510539e-03
## 85    5757.4479664  90.457294247 -8.245455e+02 -1.493878e-03  1.687310e-02
## 86    -133.4079943   4.034820196 -3.672151e+00  4.982293e-04 -3.275301e-04
## 87     335.9056218  -1.814213690 -2.203354e+01 -4.139251e-04  7.400539e-04
## 88      11.9754171  -0.371590951 -3.585944e-01  3.622461e-05  6.389082e-06
## 89     -92.1892320  -0.059319278  1.340935e+01 -1.847952e-04 -4.436388e-04
## 90      23.3693363  -0.811539703 -1.396215e+00 -4.770518e-05  2.191380e-04
## 91    -649.3200018   2.844479108  9.458721e+01  1.936196e-03 -7.592025e-03
## 92     260.9036626  -5.693386335  2.045662e+00 -1.494839e-04 -4.477405e-04
## 93     110.5551000   1.021048610 -1.126412e+01  1.233100e-06  4.055399e-05
## 94    -166.0385889   0.003844257  1.503542e+01  5.010622e-04 -8.740197e-04
## 95      68.9757628  -1.759318117  8.396497e-01  4.976831e-05 -1.725677e-04
## 96      19.6018061  -1.339694194  8.120366e-02  9.899803e-05  9.414522e-05
## 97     327.5798134  -7.178099754  2.149754e+00 -1.663746e-04 -4.829807e-04
## 98     276.3718119  -6.137758214 -5.346413e+00 -1.421445e-05  1.781608e-04
## 99    -927.0420400   9.358463331  5.281601e+01  1.019251e-03 -2.176764e-03
## 100     49.8803077  -1.087003275 -9.470440e-01  8.577011e-06  1.083237e-05
## 101   -110.7974609   1.690442721  2.041445e-01  7.115070e-05  2.796440e-04
## 102  -4401.8535090  -0.950003379  3.970242e+02  4.348645e-03 -1.196987e-02
```

```
covratio(M2)
```

```
##         1         2         3         4         5         6         7 
## 1.0666224 1.1059005 1.0702372 1.0710685 1.0310038 1.0674792 1.1012131 
##         8         9        10        11        12        13        14 
## 1.0702890 1.3467619 1.1226449 1.0911187 1.1615340 1.0864323 1.0756460 
##        15        16        17        18        19        20        21 
## 1.2505067 1.0572150 1.0587328 1.1123119 1.0601708 1.0820817 1.2739277 
##        22        23        24        25        26        27        28 
## 1.0397827 1.0645314 1.0706289 1.0654369 1.0919907 1.0231513 1.0679089 
##        29        30        31        32        33        34        35 
## 1.0182776 1.0826968 1.0595375 1.0521093 1.0541730 1.0713098 1.0690419 
##        36        37        38        39        40        41        42 
## 1.0721582 1.0693799 0.6630213 1.0443213 1.0745297 1.0288236 1.0005814 
##        43        44        45        46        47        48        49 
## 1.0325504 1.0585543 1.0671550 1.0558629 1.1112711 1.1056822 0.9911092 
##        50        51        52        53        54        55        56 
## 1.0648846 1.0666818 0.8621676 1.0737421 1.0551394 1.0684877 1.0634387 
##        57        58        59        60        61        62        63 
## 1.0244180 1.0713607 1.2788963 1.0719271 1.0686089 0.9606044 1.0673606 
##        64        65        66        67        68        69        70 
## 0.9707150 1.0648013 1.0742445 1.0384399 1.0942271 1.0368104 1.0607725 
##        71        72        73        74        75        76        77 
## 1.0655539 1.0633975 1.0018020 1.0691312 1.0798585 1.0806151 1.1853654 
##        78        79        80        81        82        83        84 
## 1.0824365 1.0480625 1.0752445 1.0325091 1.0382715 1.0739267 0.7015716 
##        85        86        87        88        89        90        91 
## 0.9742316 1.0535013 1.0360986 1.0908757 1.0729400 1.0727354 0.5385162 
##        92        93        94        95        96        97        98 
## 1.0444394 1.1084142 1.1161399 1.1049987 1.0779844 0.9744366 1.1884745 
##        99       100       101       102 
## 1.0044974 1.2137020 1.0819317 0.5571804
```

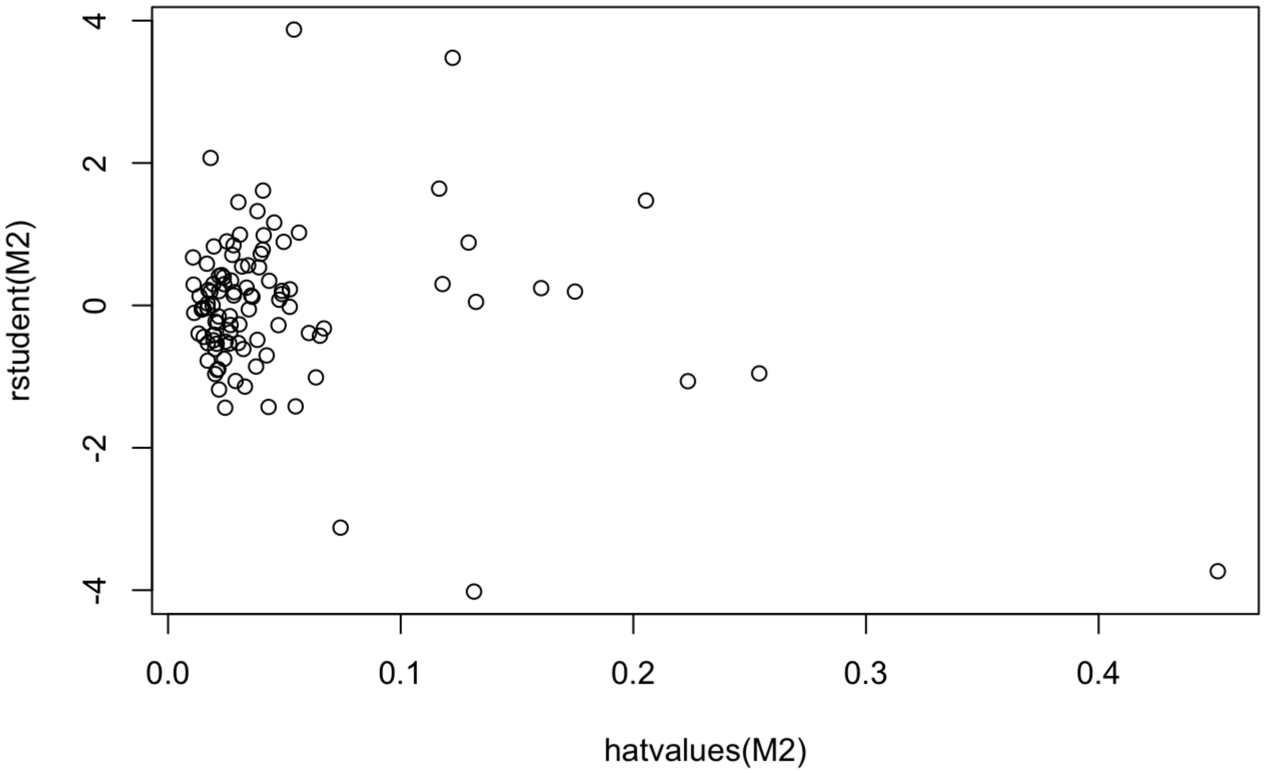
```
dffits(M2)
```

```
##             1             2             3             4             5 
## -0.2634430059  0.0361301081  0.0618534001  0.0989758355 -0.1332303423 
##             6             7             8             9            10 
##  0.1472650224 -0.0619568350 -0.0932900706 -0.5574959746 -0.0867005783 
##            11            12            13            14            15 
##  0.0265929967  0.3398822121  0.0465334074  0.0582461795  0.1059399030 
##            16            17            18            19            20 
##  0.2495413360 -0.0552862597 -0.0985476301  0.0307354192  0.0317567114 
##            21            22            23            24            25 
##  0.0894089774  0.0701627690 -0.0112099683  0.0260089976 -0.0811257672 
##            26            27            28            29            30 
##  0.0232112180 -0.1835408981  0.0653631986 -0.2107902925  0.0236627980 
##            31            32            33            34            35 
## -0.0787521177  0.0763053025 -0.0879792715 -0.0049155590  0.0422041700 
##            36            37            38            39            40 
## -0.0330610594 -0.0050466937  1.2982583906  0.1436582821  0.0299478825 
##            41            42            43            44            45 
##  0.2544489724  0.2645082974 -0.1330624389 -0.0455484313  0.0620153680 
##            46            47            48            49            50 
## -0.0700684305 -0.0047277692  0.0178037827 -0.3031372840 -0.0588459283 
##            51            52            53            54            55 
##  0.0152472736  0.2825031761  0.0001930679  0.1195183369 -0.0071566733 
##            56            57            58            59            60 
##  0.1613332709 -0.1386393564  0.0034318498 -0.5711453963 -0.1477628795 
##            61            62            63            64            65 
## -0.0049490075  0.3326169700 -0.1120136670 -0.2279584984 -0.0580379353 
##            66            67            68            69            70 
##  0.0468113824 -0.1020983248  0.0735518389  0.1172607448 -0.0692127995 
##            71            72            73            74            75 
## -0.0881622024  0.2041438234 -0.1767480685  0.0287049852  0.1077997141 
##            76            77            78            79            80 
## -0.0248914192  0.7487623346 -0.0473428421 -0.1179461735 -0.0232144530 
##            81            82            83            84            85 
##  0.1774608615  0.5956523825 -0.0630258998 -0.8832273526 -3.3856179801 
##            86            87            88            89            90 
## -0.1702810151  0.1446834324 -0.0106075781  0.1061278377 -0.0361591599 
##            91            92            93            94            95 
##  0.9266752590  0.2039711450  0.0530495395 -0.1125806637  0.0463435839 
##            96            97            98            99           100 
## -0.0457289899  0.2560096459  0.1096417061 -0.3415864978  0.0193359514 
##           101           102 
## -0.0966942184 -1.5640269321
```

```
cooks.distance(M2)
```

```
##            1            2            3            4            5 
## 1.387721e-02 2.637268e-04 7.718831e-04 1.973525e-03 3.556584e-03 
##            6            7            8            9           10 
## 4.358840e-03 7.751049e-04 1.753632e-03 6.221661e-02 1.517403e-03 
##           11           12           13           14           15 
## 1.428824e-04 2.315681e-02 4.373002e-04 6.847220e-04 2.266647e-03 
##           16           17           18           19           20 
## 1.244849e-02 6.164226e-04 1.959484e-03 1.907323e-04 2.037257e-04 
##           21           22           23           24           25 
## 1.614813e-03 9.901366e-04 2.539155e-05 1.366509e-04 1.326406e-03 
##           26           27           28           29           30 
## 1.088583e-04 6.728621e-03 8.617444e-04 8.859045e-03 1.131294e-04 
##           31           32           33           34           35 
## 1.249519e-03 1.172439e-03 1.558111e-03 4.882812e-06 3.596126e-04 
##           36           37           38           39           40 
## 2.207631e-04 5.146796e-06 3.024995e-01 4.139797e-03 1.811681e-04 
##           41           42           43           44           45 
## 1.290165e-02 1.388535e-02 3.548356e-03 4.185739e-04 7.758044e-04 
##           46           47           48           49           50 
## 9.892349e-04 4.516908e-06 6.405109e-05 1.818434e-02 6.985344e-04 
##           51           52           53           54           55 
## 4.697191e-05 1.543810e-02 7.532700e-09 2.871636e-03 1.034992e-05 
##           56           57           58           59           60 
## 5.226508e-03 3.846982e-03 2.380038e-06 6.515161e-02 4.389692e-03 
##           61           62           63           64           65 
## 4.949476e-06 2.176755e-02 2.525678e-03 1.028018e-02 6.795031e-04 
##           66           67           68           69           70 
## 4.424178e-04 2.093401e-03 1.091891e-03 2.758980e-03 9.656410e-04 
##           71           72           73           74           75 
## 1.566013e-03 8.352439e-03 6.222610e-03 1.664308e-04 2.341398e-03 
##           76           77           78           79           80 
## 1.251777e-04 1.107941e-01 4.526038e-04 2.794838e-03 1.088774e-04 
##           81           82           83           84           85 
## 6.299248e-03 6.974540e-02 8.015187e-04 1.431195e-01 2.022639e+00 
##           86           87           88           89           90 
## 5.814865e-03 4.194951e-03 2.273782e-05 2.268621e-03 2.640556e-04 
##           91           92           93           94           95 
## 1.500696e-01 8.323362e-03 5.684121e-04 2.556453e-03 4.338315e-04 
##           96           97           98           99          100 
## 4.222514e-04 1.296071e-02 2.427033e-03 2.309534e-02 7.555279e-05 
##          101          102 
## 1.884832e-03 4.231273e-01
```

```
# plot of deleted studentized residuals vs hat values
plot(rstudent(M2)~hatvalues(M2))
```




```
rstudent(M2)
```

```
##           1           2           3           4           5 
## -1.011248383  0.159290150  0.395239466  0.545921886 -0.906749015 
##           6           7           8           9          10 
##  0.723069227 -0.277603207 -0.528748496 -0.955137647 -0.323599900 
##          11          12          13          14          15 
##  0.138150535  0.882451634  0.249050737  0.349133778  0.242404925 
##          16          17          18          19          20 
##  1.021901080 -0.442872436 -0.388151732  0.291653765  0.185567309 
##          21          22          23          24          25 
##  0.194193488  0.673764806 -0.105171511  0.191196128 -0.509259166 
##          26          27          28          29          30 
##  0.119663433 -1.061746470  0.425592608 -1.140490437  0.138812289 
##          31          32          33          34          35 
## -0.538968449  0.585805613 -0.612333307 -0.037384582  0.299773360 
##          36          37          38          39          40 
## -0.229191037 -0.040576387  3.477560544  0.844279442  0.200092501 
##          41          42          43          44          45 
##  1.164012804  1.323439050 -0.895696625 -0.394995006  0.414600837 
##          46          47          48          49          50 
## -0.531452468 -0.020126309  0.079464767 -1.426673416 -0.414223094 
##          51          52          53          54          55 
##  0.130196346  2.071065929  0.001381733  0.709321675 -0.058989514 
##          56          57          58          59          60 
##  0.783292398 -0.963954209  0.026092672 -1.064756714 -0.702550770 
##          61          62          63          64          65 
## -0.040766229  1.612742554 -0.612627281 -1.436888332 -0.410854676 
##          66          67          68          69          70 
##  0.297737827 -0.775932521  0.345066744  0.827562880 -0.490629636 
##          71          72          73          74          75 
## -0.536474317  0.892621698 -1.181273327  0.216244453  0.534484710 
##          76          77          78          79          80 
## -0.150746433  1.472651771 -0.266392411 -0.750631566 -0.155665776 
##          81          82          83          84          85 
##  0.994007124  1.640037930 -0.380633906 -3.121237593 -3.733749573 
##          86          87          88          89          90 
## -0.858738641  0.899036252 -0.055936944  0.562152056 -0.245126575 
##          91          92          93          94          95 
##  3.874339986  0.985226270  0.225710184 -0.426012856  0.204246050 
##          96          97          98          99         100 
## -0.275266652  1.450439545  0.299792171 -1.418411117  0.049503941 
##         101         102 
## -0.484079295 -4.019345387
```

Answer: This plot is the most important and it shows a number of points that I argue are influential points. Specifically, there are 5 points that are outside of the (-3,3) studentized residuals band, suggesting they are influential points. One of those points also has a high hat value close to 0.5 so it’s definitely influential.

This makes sense, because when I looked at the normal qq plot, there were 5 points that didn’t fall well along the line. **If they were removed, I believe the model would improve.**

d) 4

Compute the standardized coefficients and discuss which predictor has the strongest influence on balance? [1 pt R code, 1 pt answer = 2 pts]

```
# lm.beta function
library(QuantPsyc)
```

```
## Loading required package: boot
```

```
##
## Attaching package: 'boot'
```

```
## The following object is masked from 'package:car':
##
##     logit
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'QuantPsyc'
```

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

```
lm.beta(M2)
```

```
##        age  education     wealth    homeval
## 0.15227001 0.08902878 0.75821456 0.11085473
```

Answer:

These standardized coefficients provide a unitless measure of the effect of the 4 predictor variables I included in my model on the response variable balance. The strongest predictor of Y is Wealth, followed by Age, then Homevalue, and finally education. This order based on the size of the standardized coefficients.

Problem 1 e)

Use the fitted regression model from d) without removal of influential points to predict the average bank balance for a specific zip code area where there is a plan to open a new branch.

Census data in that area show the following values: * median age is 34 years * median education is 13 years * median income is $64,000 * median home value is $140,000 * median wealth is 160,000. **(Note that you may not need all these values in your model). Provide predicted average bank balance and 95% confidence interval for your estimate. [2 pts R code, 2 pts answer = 4 pts]**

```
new = data.frame(age=c(34), education=c(13),wealth=c(160000), homeval=c(140000))
# compute average response value and confidence interval
predict(M2, new, interval="confidence",level=0.95)
```

```
##     fit      lwr      upr
## 1 30848 30003.24 31692.76
```

Answer: So the average bank balance for those values would be 30,848 dollars based on my M2 model. The lower bound of 30,003.24 and upper bound of 31692.76 for the 95% confidence interval.

Problem 2 [24 points]

Analytics is used in many different sports and has become popular with the Money Ball movie. The pgatour2006.csv dataset contains data about 196 tour players in 2006. The variables in the dataset are: * Player's name * PrizeMoney = average prize money per tournament

And a set of metrics that evaluate the quality of a player's game. * DrivingAccuracy = percent of times a player is able to hit the fairway with his tee shot * GIR = percent of time a player was able to hit the green within two or less than par (Greens in Regulation) * BirdieConversion = percentage of times a player makes a birdie or better after hitting the green in regulation * PuttingAverage = putting performance on those holes where the green was hit in regulation. * PuttsPerRound = average number of putts per round (shots played on the green)

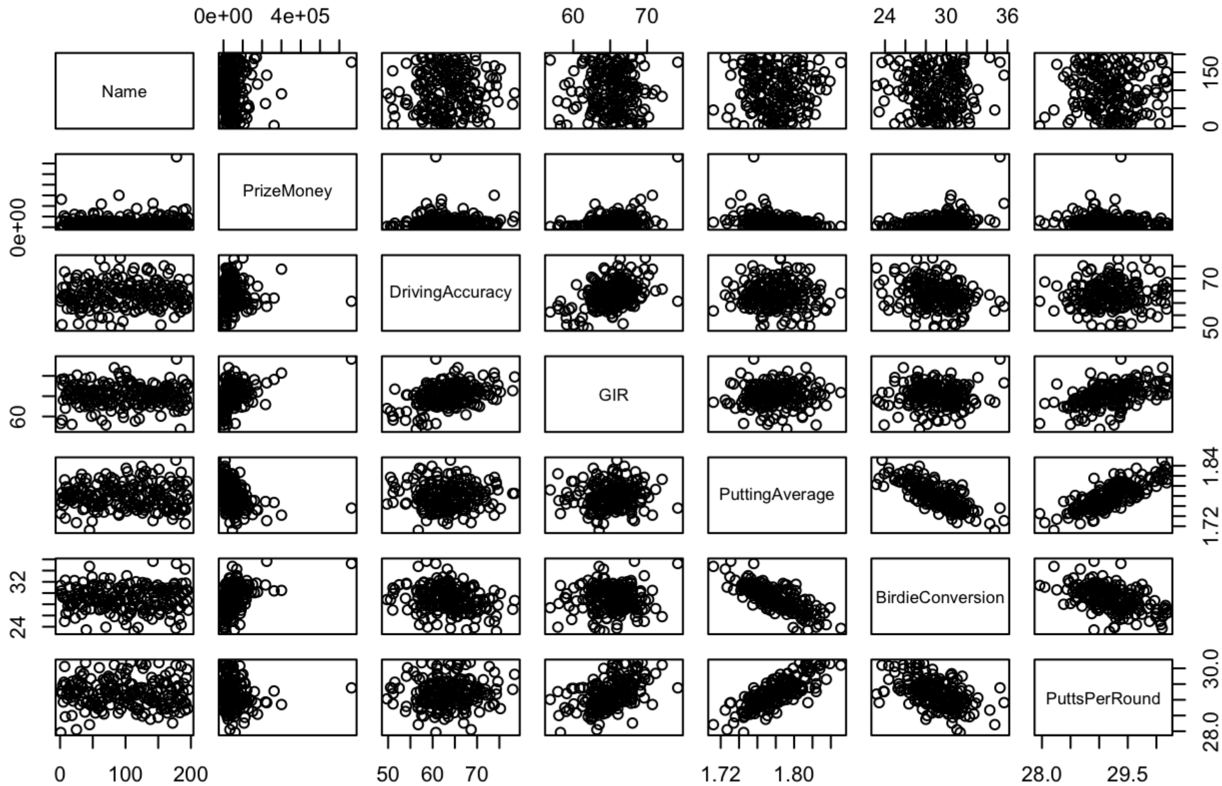
You are asked to build a model for PrizeMoney using the remaining predictors, and to evaluate the relative importance of each different aspects of a player's game on the average prize money. For the non golfers in the class, you can refer to this page for an explanation of the terms: http://en.wikipedia.org/wiki/Glossary_of_golf (http://en.wikipedia.org/wiki/Glossary_of_golf)

Problem 2 a)

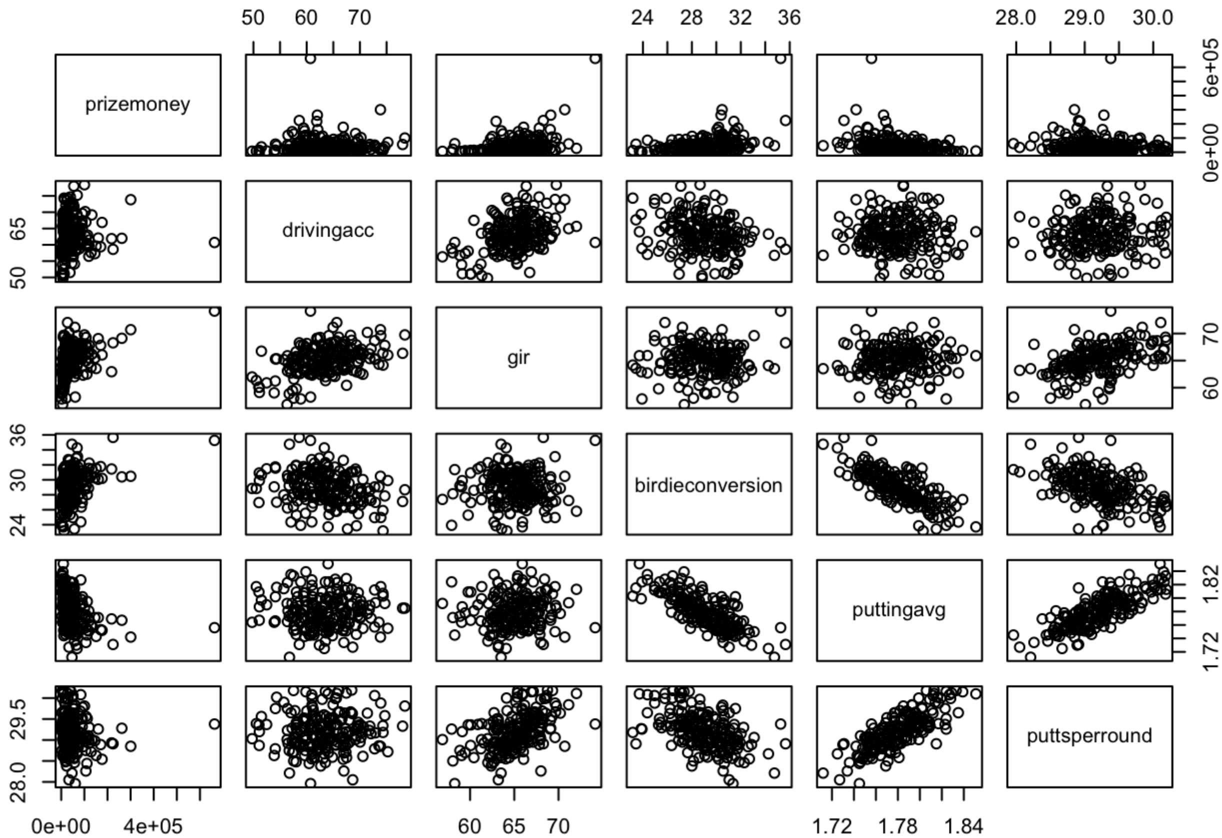
Create scatterplots to visualize the associations between PrizeMoney and the other five variables. Discuss the patterns displayed by the scatterplot. Do the associations appear to be linear? (you can create scatterplots or a matrix plot) [1 pt R code, 1 pt scatterplots, 2 pts answer = 4 pts]

```
## load in the data from file
pgadata=read.csv("pgatour2006_small.csv", header=T)

## get the variables
name=pgadata$Name
prizemoney=pgadata$PrizeMoney

drivingacc=pgadata$DrivingAccuracy
gir=pgadata$GIR
birdieconversion=pgadata$BirdieConversion
puttingavg=pgadata$PuttingAverage
puttsperround=pgadata$PuttsPerRound

# create scatterplot matrix for quantitative variables
plot(pgadata, main="Simple Scatterplot Matrix")
```

Simple Scatterplot Matrix

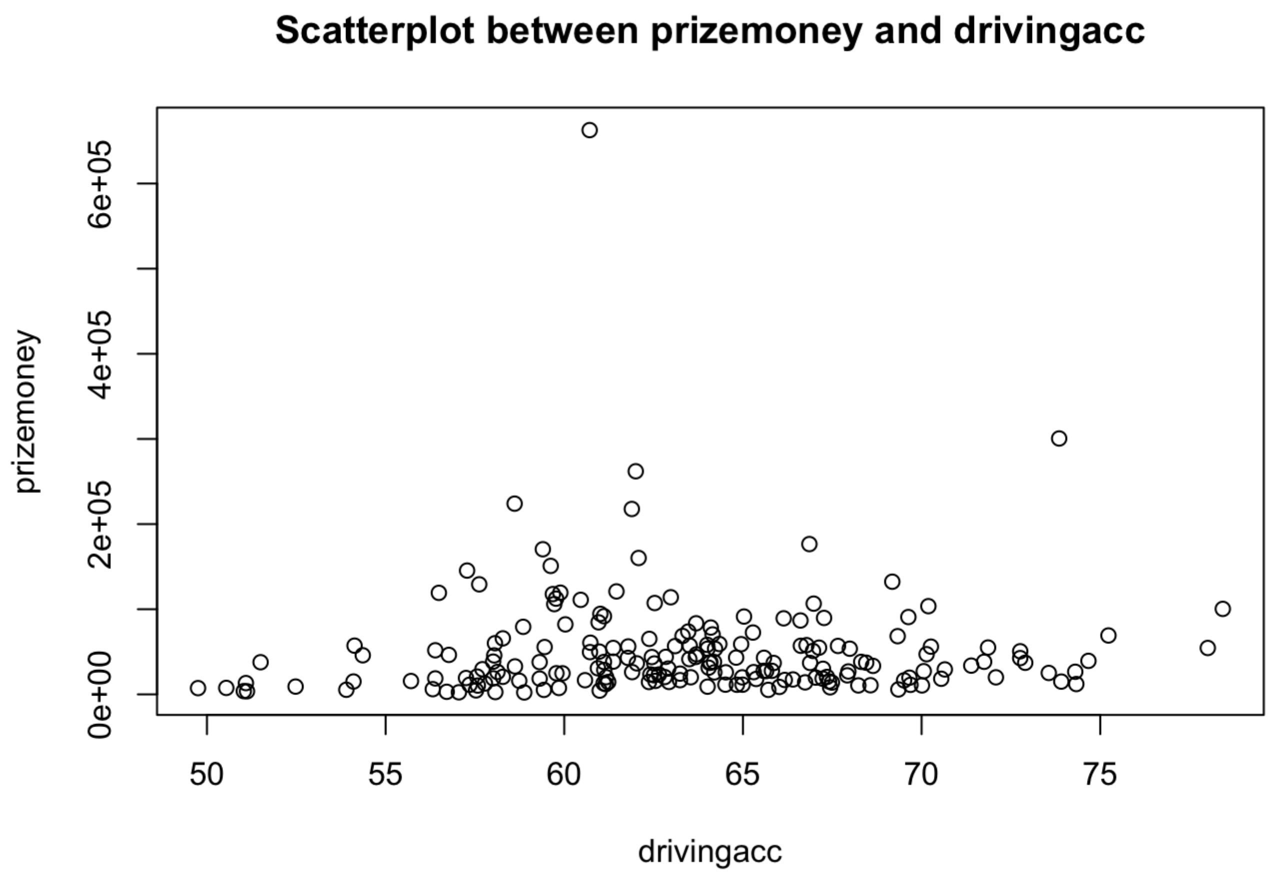



```
#prizemoney pairs
pairs(prizemoney ~ drivingacc+gir+birdieconversion+puttingavg+puttsperround)
```



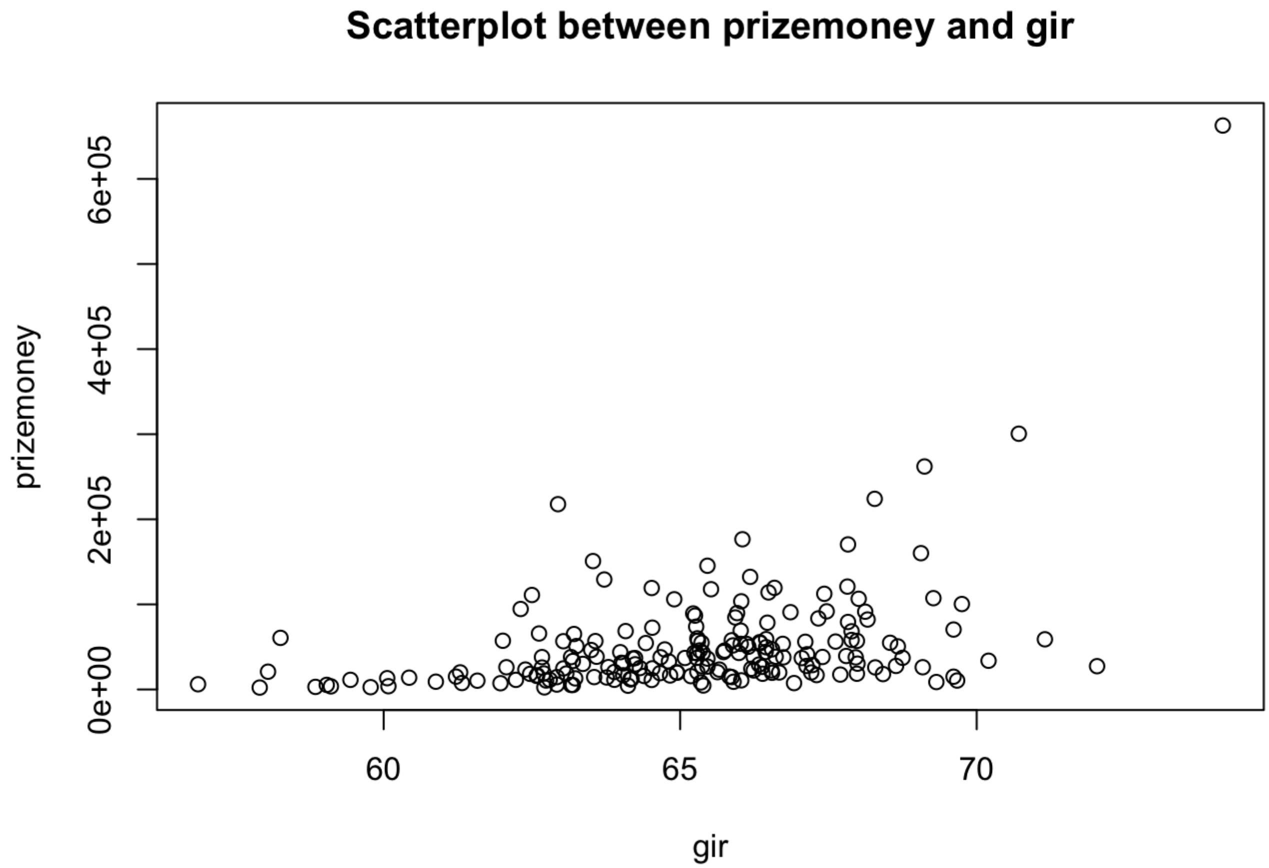
Answer: The associations do not appear to be linear between the 5 variables and prizemoney. In order to apply a regression model, we must first transform the data.

```
# 1 scatterplot between prizemoney and drivingacc
plot(drivingacc, prizemoney, main="Scatterplot between prizemoney and drivingacc", xlab=
"drivingacc", ylab="prizemoney")
```



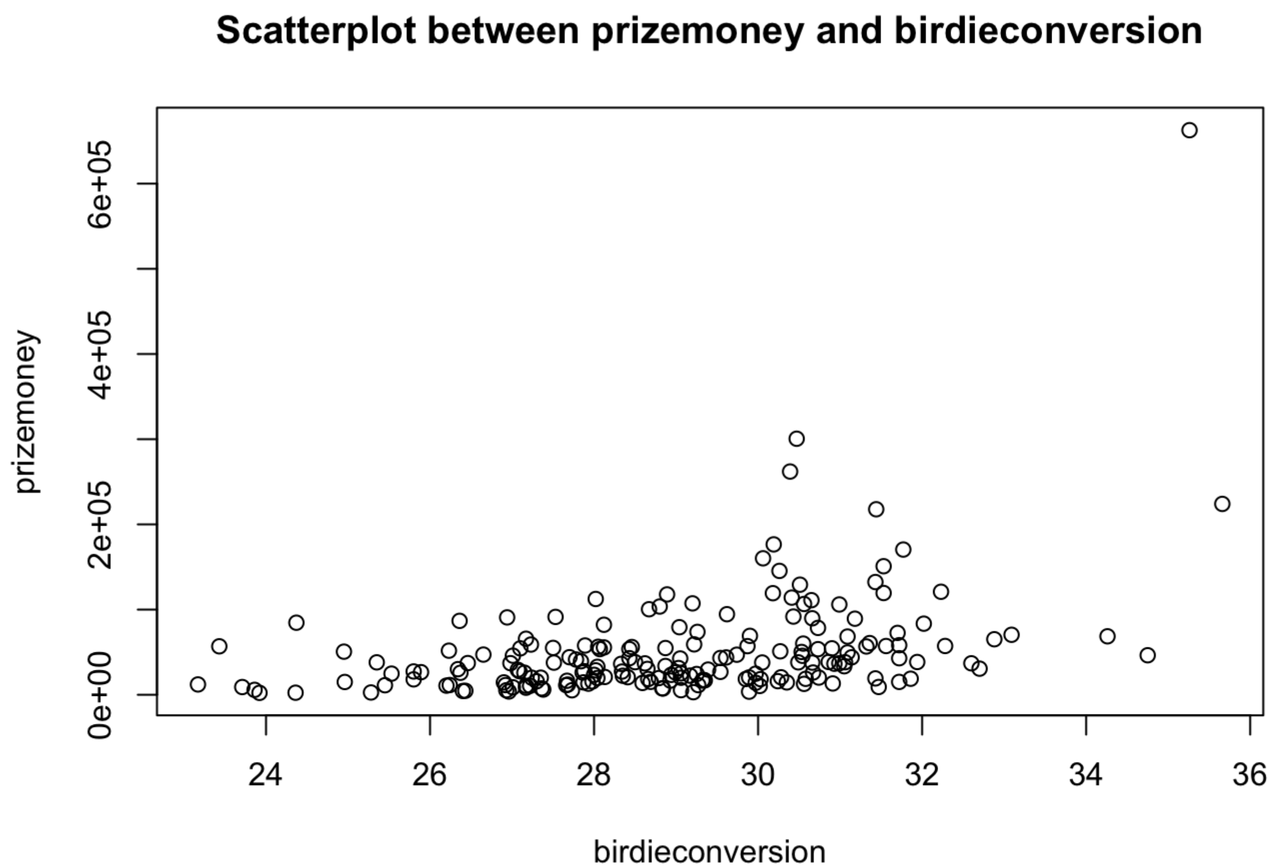
Answer: Not linear

```
# 2 scatterplot between prizemoney and gir
plot(gir, prizemoney, main="Scatterplot between prizemoney and gir", xlab="gir", ylab="p
rizemoney")
```



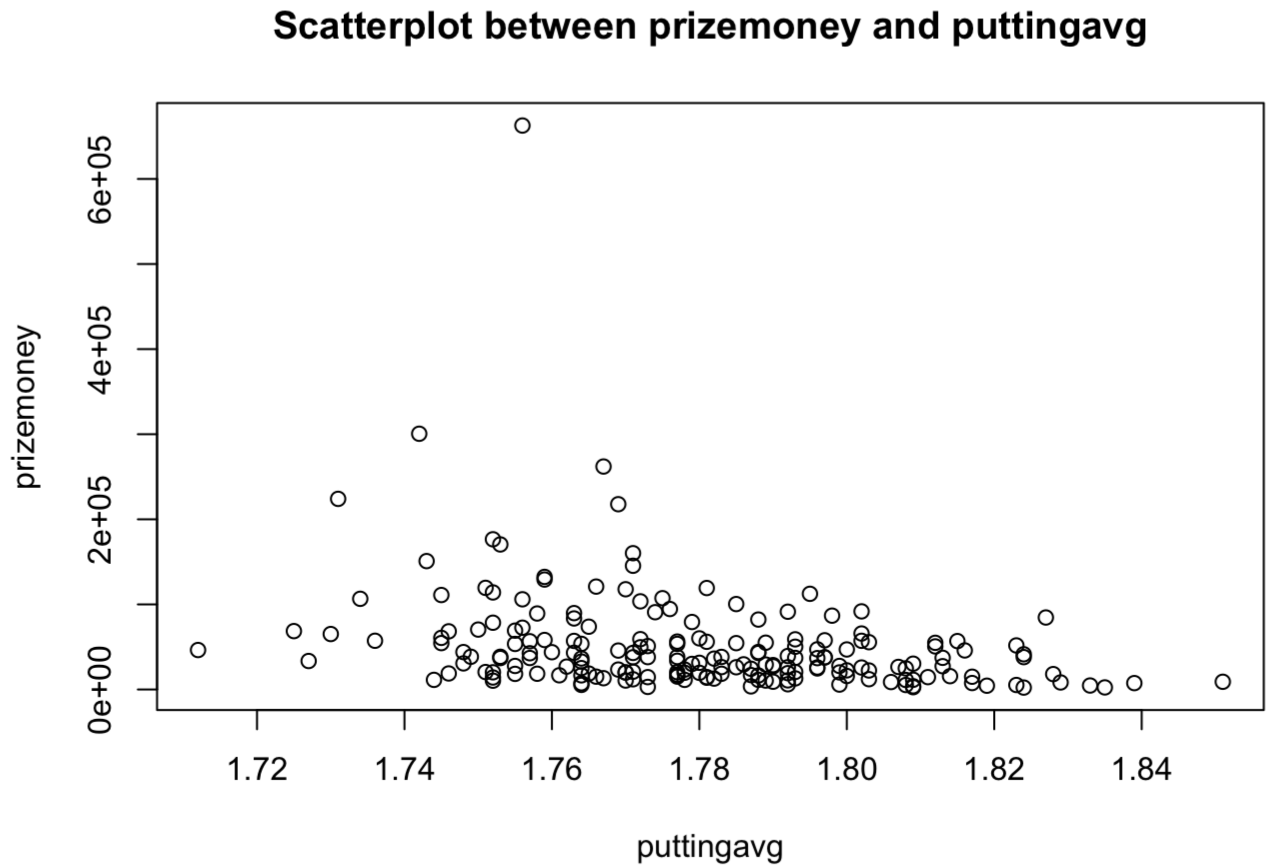
Answer: Not linear

```
# 3 scatterplot between prizemoney and birdieconversion
plot(birdieconversion, prizemoney, main="Scatterplot between prizemoney and birdieconver
sion", xlab="birdieconversion", ylab="prizemoney")
```



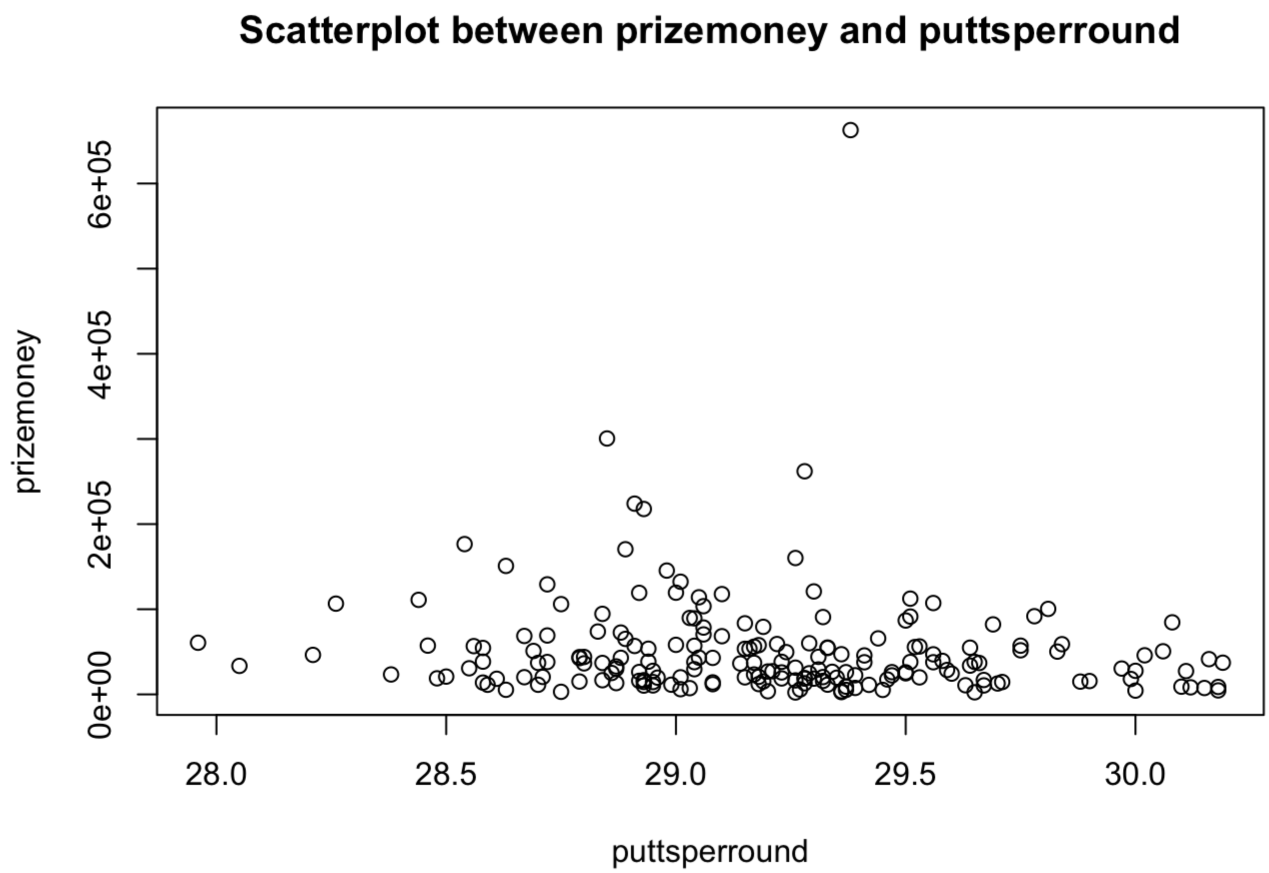
Answer: Not linear

```
# 4 scatterplot between prizemoney and puttingavg
plot(puttingavg, prizemoney, main="Scatterplot between prizemoney and puttingavg", xlab=
"puttingavg", ylab="prizemoney")
```



Answer: Not linear

```
# 5 scatterplot between prizemoney and puttsperround
plot(puttsperround, prizemoney, main="Scatterplot between prizemoney and puttsperround",
 xlab="puttsperround", ylab="prizemoney")
```



Answer: Not linear

Problem 2 b)

Analyze distribution of PrizeMoney, and discuss if the distribution is symmetric or skewed. [1 pt R code, 1 pt answer = 2 pts]

```
# plot the histogram of account prizemoney
hist(prizemoney)

# compute descriptive statistics
library(psych)
```

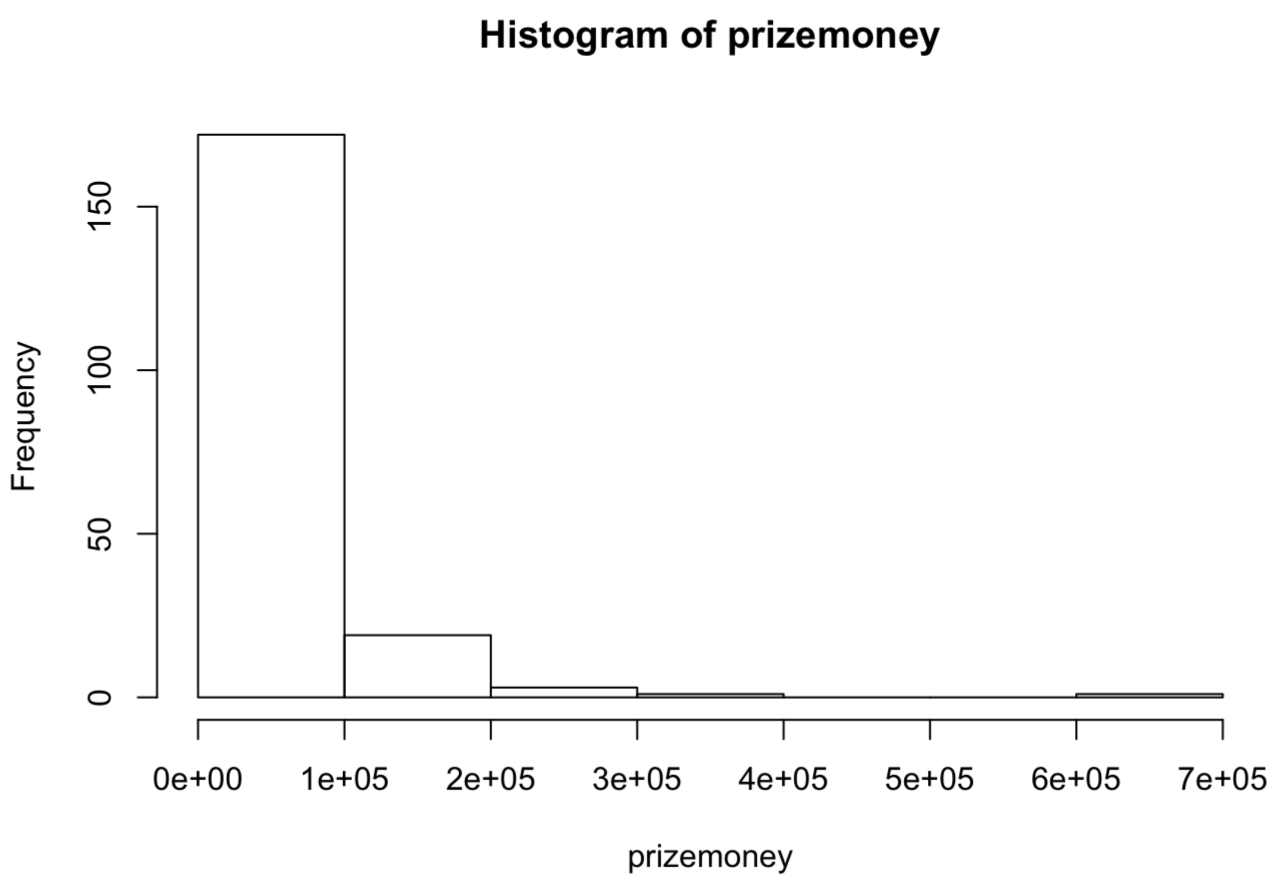
```
## Warning: package 'psych' was built under R version 3.5.2
```

```
##
## Attaching package: 'psych'
```

```
## The following object is masked _by_ '.GlobalEnv':
##
##     income
```

```
## The following object is masked from 'package:boot':
##
##     logit
```

```
## The following object is masked from 'package:car':
##
##     logit
```




```
describe(prizemoney)
```

```
##    vars   n     mean       sd  median  trimmed      mad  min    max  range
## X1    1 196 50891.17 63902.95 36644.5 40027.22 30153.12 2240 662771 660531
##    skew kurtosis     se
## X1 5.29    42.57 4564.5
```

```
# get the quantile
quantile(prizemoney)
```

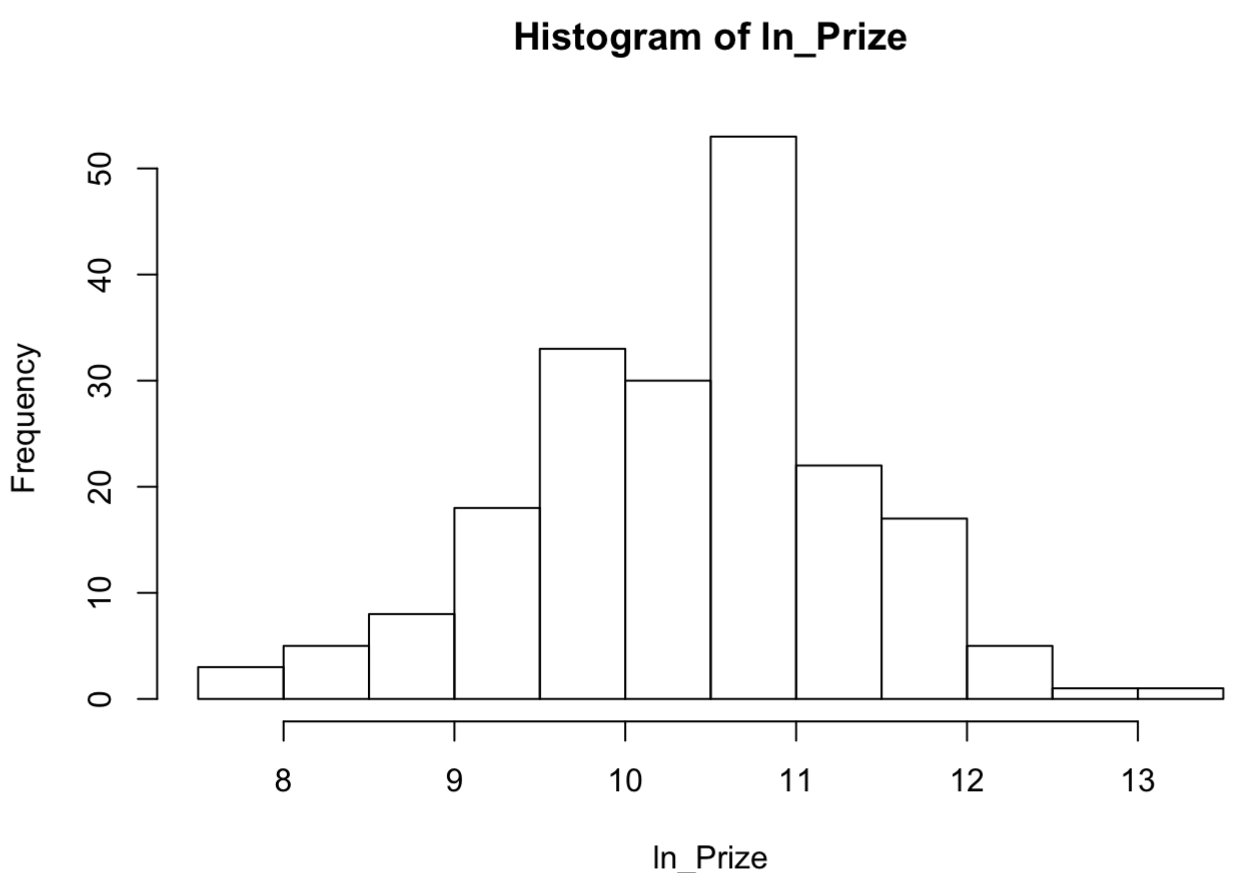
```
##        0%       25%       50%       75%      100%
##   2240.00  17368.75  36644.50  57915.25 662771.00
```

Answer: The data is clearly skewed and not symmetric. It is skewed to the 'right'. This shows as a long tail to the right, a mean greater than the median, a massive standard deviation of 63902, skew of 5.29, kurtosis of 42.57. It is clear that the data must be transformed.

Problem 2 c)

Apply a log transformation to PrizeMoney and compute the new variable ln_Prize=log(PrizeMoney). Analyze distribution of ln_Prize, and discuss if the distribution is symmetric or skewed. [2 pts R code, 1 pt answer = 3 pts]

```
#transform
ln_Prize=log(prizemoney)

# plot the histogram of account prizemoney
hist(ln_Prize)
```




```
# compute descriptive statistics
library(psych)
describe(ln_Prize)
```

```
##    vars   n  mean   sd median trimmed  mad  min  max range skew kurtosis
## X1    1 196 10.38 0.98  10.51   10.41 0.93 7.71 13.4  5.69 -0.2     0.18
##      se
## X1 0.07
```

```
# get the quantile
quantile(ln_Prize)
```

```
##        0%       25%       50%       75%      100%
##  7.714231  9.762319 10.509001 10.966732 13.404185
```

Answer: The distribution is symmetrical now. The mean of 10.38 is close to median 10.51, skew is only -0.2, and standard deviation 0.98, for a range of 5.69. The transformation worked.

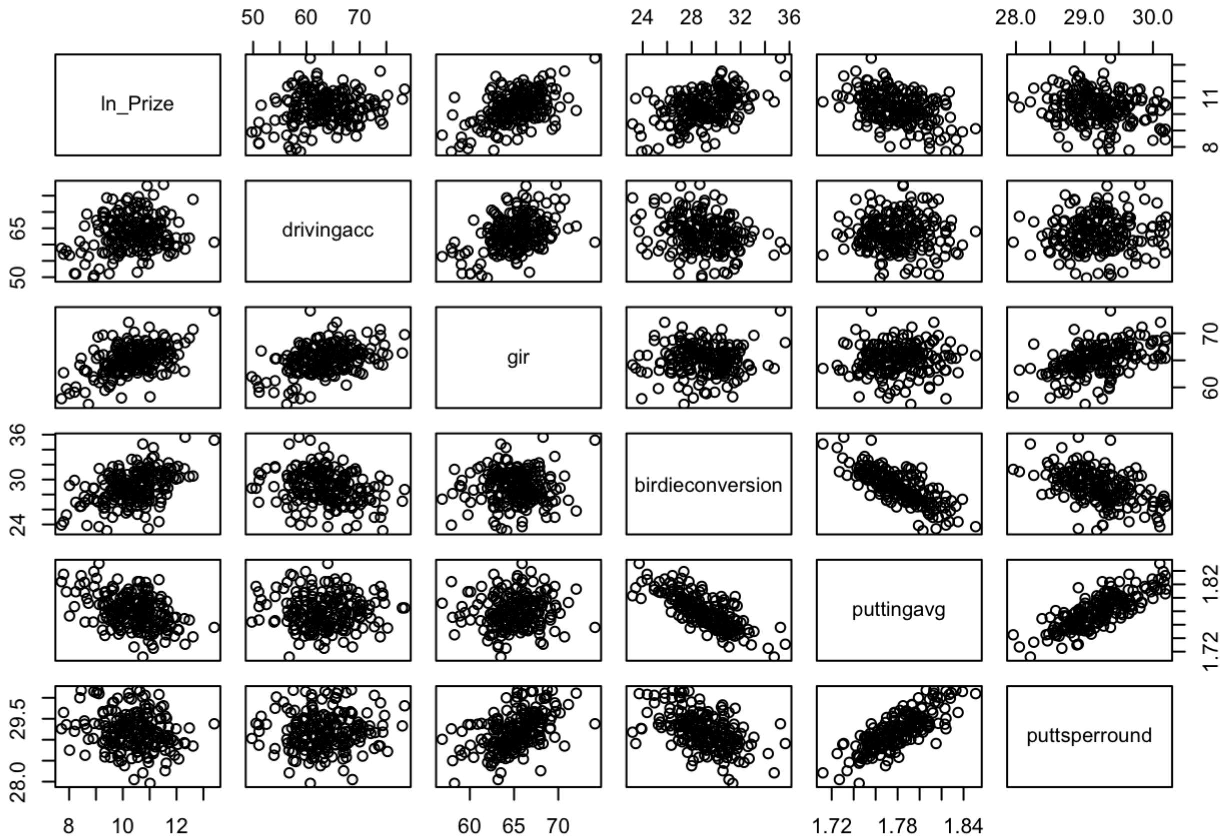
Problem 2 d)

Fit a regression model of ln_Prize using the remaining predictors in your dataset. Apply your knowledge of regression analysis to define a valid model to predict ln_Prize. Hint: use scatterplots and correlation [3 pts R code, 1 pt answer = 4 pts]

```
fitpga <- lm(ln_Prize ~ drivingacc+gir+birdieconversion+puttingavg+puttsperround)
summary(fitpga)
```

```
##
## Call:
## lm(formula = ln_Prize ~ drivingacc + gir + birdieconversion +
##     puttingavg + puttsperround)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.55696 -0.51250 -0.08005  0.45090  2.11898
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       8.2410192  7.1611241   1.151 0.251261
## drivingacc       -0.0007584  0.0116109  -0.065 0.947992
## gir               0.2687898  0.0287938   9.335  < 2e-16 ***
## birdieconversion  0.1523018  0.0408329   3.730 0.000253 ***
## puttingavg        8.7467774  5.3734220   1.628 0.105228
## puttsperround    -1.2094847  0.2672761  -4.525 1.06e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6725 on 190 degrees of freedom
## Multiple R-squared:  0.5414, Adjusted R-squared:  0.5293
## F-statistic: 44.86 on 5 and 190 DF,  p-value: < 2.2e-16
```

```
#prizemoney pairs
pairs(ln_Prize ~ drivingacc+gir+birdieconversion+puttingavg+puttsperround)
```



*** If necessary remove not significant variables. Remember to remove one variable at a time (variable with largest p-value is removed first) and refit the model, until all variables are significant. [2 pts R code, 1 pt answer = 3 pts]**

```
#id insignficant vars
summary(fitpga)
```

```
##
## Call:
## lm(formula = ln_Prize ~ drivingacc + gir + birdieconversion +
##     puttingavg + puttsperround)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.55696 -0.51250 -0.08005  0.45090  2.11898
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       8.2410192  7.1611241   1.151 0.251261
## drivingacc       -0.0007584  0.0116109  -0.065 0.947992
## gir               0.2687898  0.0287938   9.335  < 2e-16 ***
## birdieconversion  0.1523018  0.0408329   3.730 0.000253 ***
## puttingavg        8.7467774  5.3734220   1.628 0.105228
## puttsperround    -1.2094847  0.2672761  -4.525 1.06e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6725 on 190 degrees of freedom
## Multiple R-squared:  0.5414, Adjusted R-squared:  0.5293
## F-statistic: 44.86 on 5 and 190 DF,  p-value: < 2.2e-16
```

```
vif(fitpga)
```

```
##       drivingacc              gir birdieconversion       puttingavg
##         1.703301         2.649566         3.500528         7.613214
##    puttsperround
##         6.009842
```

Answer: No multicollinearity detected. Driving Accuracy and Putting Average are identified as not signficant. First I will remove drivingacc as it has the highest p value at 0.94. Currently, gir, birdie conversion, and putts per round are significant at the 5% significant level.

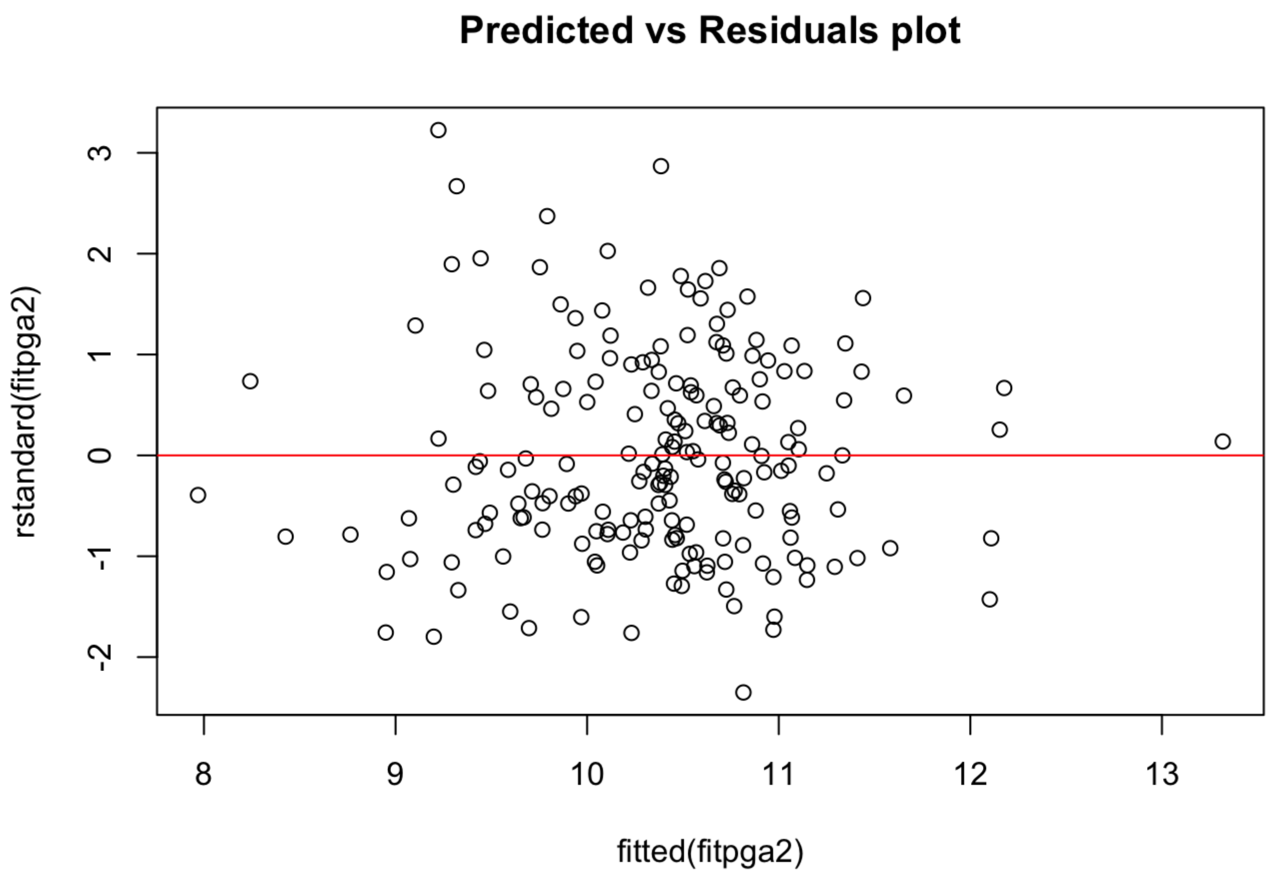
```
fitpga2 <- lm(ln_Prize ~ gir+birdieconversion+puttingavg+puttsperround)
#id insignficant vars
summary(fitpga2)
```

```
##
## Call:
## lm(formula = ln_Prize ~ gir + birdieconversion + puttingavg +
##     puttsperround)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.55608 -0.51122 -0.08109  0.45250  2.12227
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       8.02738    6.35383   1.263   0.2080
## gir               0.26791    0.02536  10.563  < 2e-16 ***
## birdieconversion  0.15360    0.03561   4.314 2.57e-05 ***
## puttingavg        8.81065    5.26991   1.672   0.0962 .
## puttsperround    -1.20702    0.26391  -4.574 8.61e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6707 on 191 degrees of freedom
## Multiple R-squared:  0.5414, Adjusted R-squared:  0.5318
## F-statistic: 56.37 on 4 and 191 DF,  p-value: < 2.2e-16
```

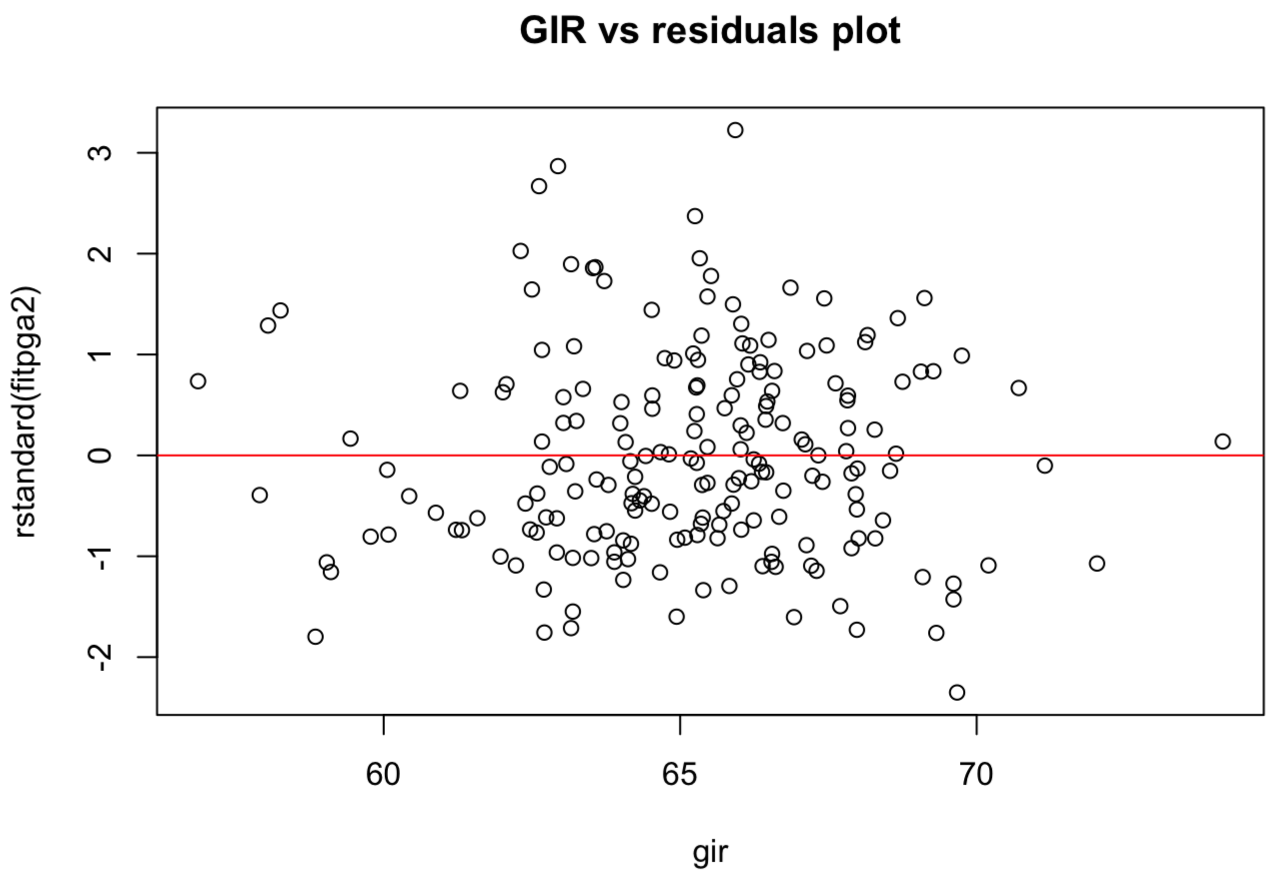
Answer: Now all the variables are significant at the 5% significance level. So my model2 will include: gir, birdieconversion, puttingaverage, and puttsperround. The F-Statistic of 56.37 and p-value of 2.2e-16 < 0.01 suggest a reasonably goodness of model fit. The Adj-R2 is still pretty low though at 0.53.

*** Analyze residual plots to check if the regression model is valid for your data. [1 pt R code, 1 pt answer = 2 pts]**

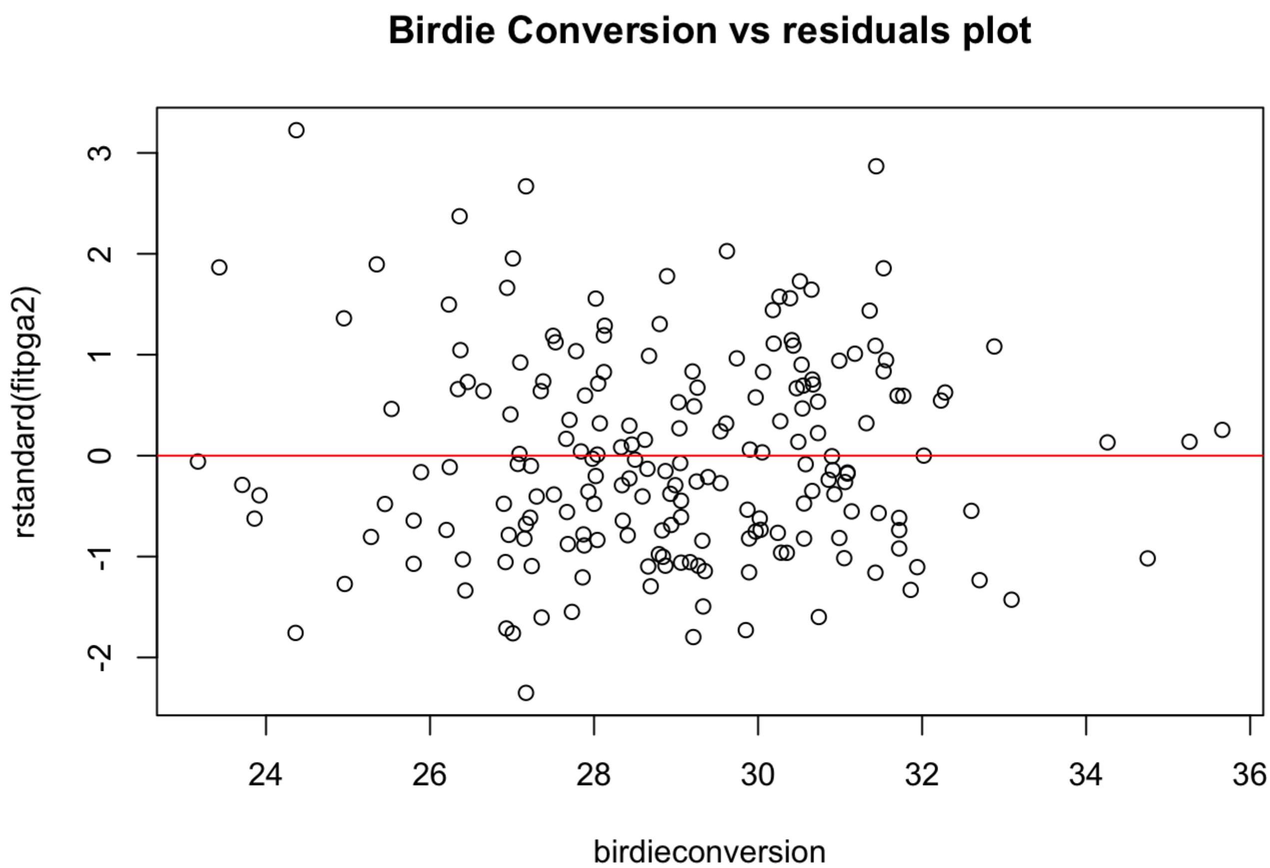
```
#residual plots
#Plot residuals vs predicted values
plot( fitted(fitpga2), rstandard(fitpga2), main="Predicted vs Residuals plot")
abline(a=0, b=0, col='red') #add zero line
```



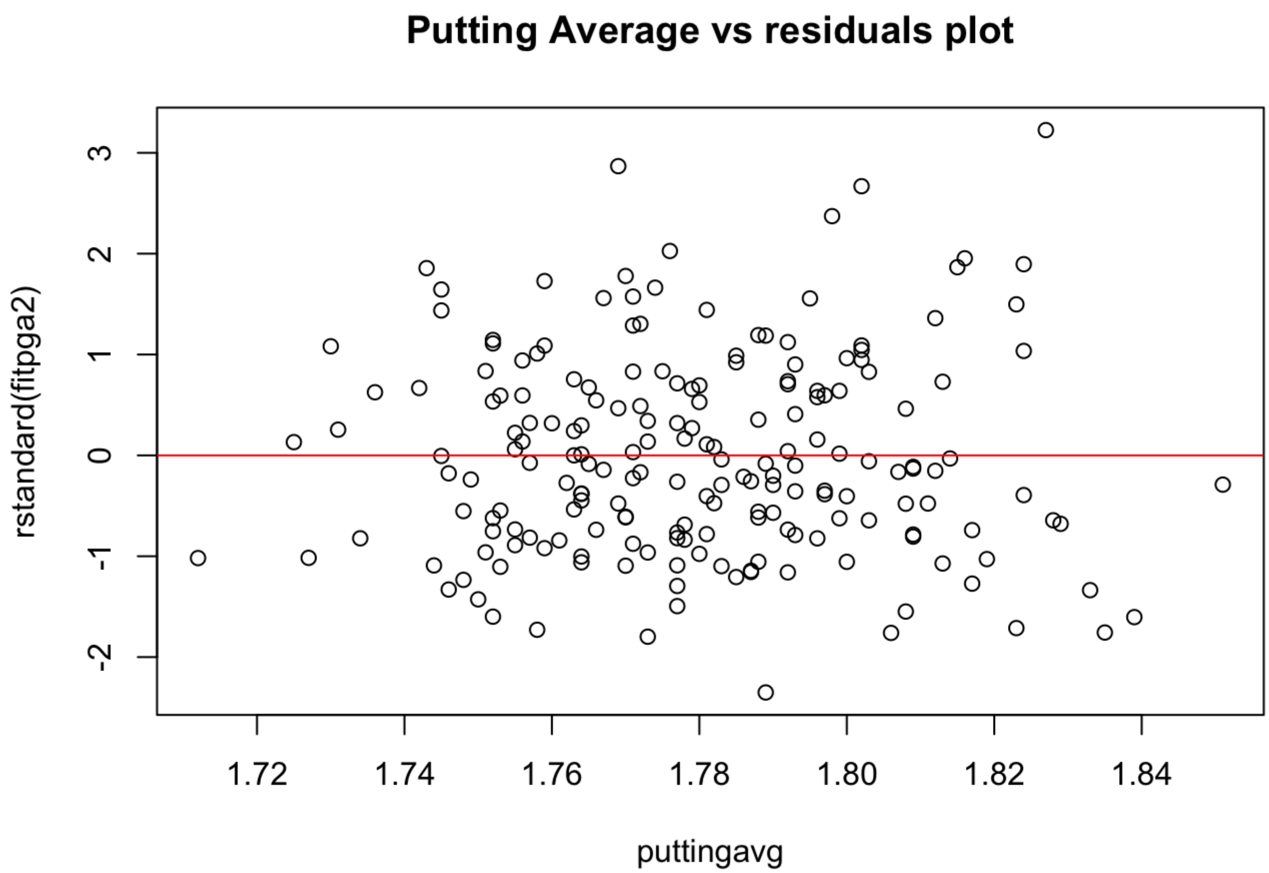

```
#Plot residuals vs each x-variable:
plot(gir, rstandard(fitpga2), main="GIR vs residuals plot")
abline(a=0, b=0,col='red')
```



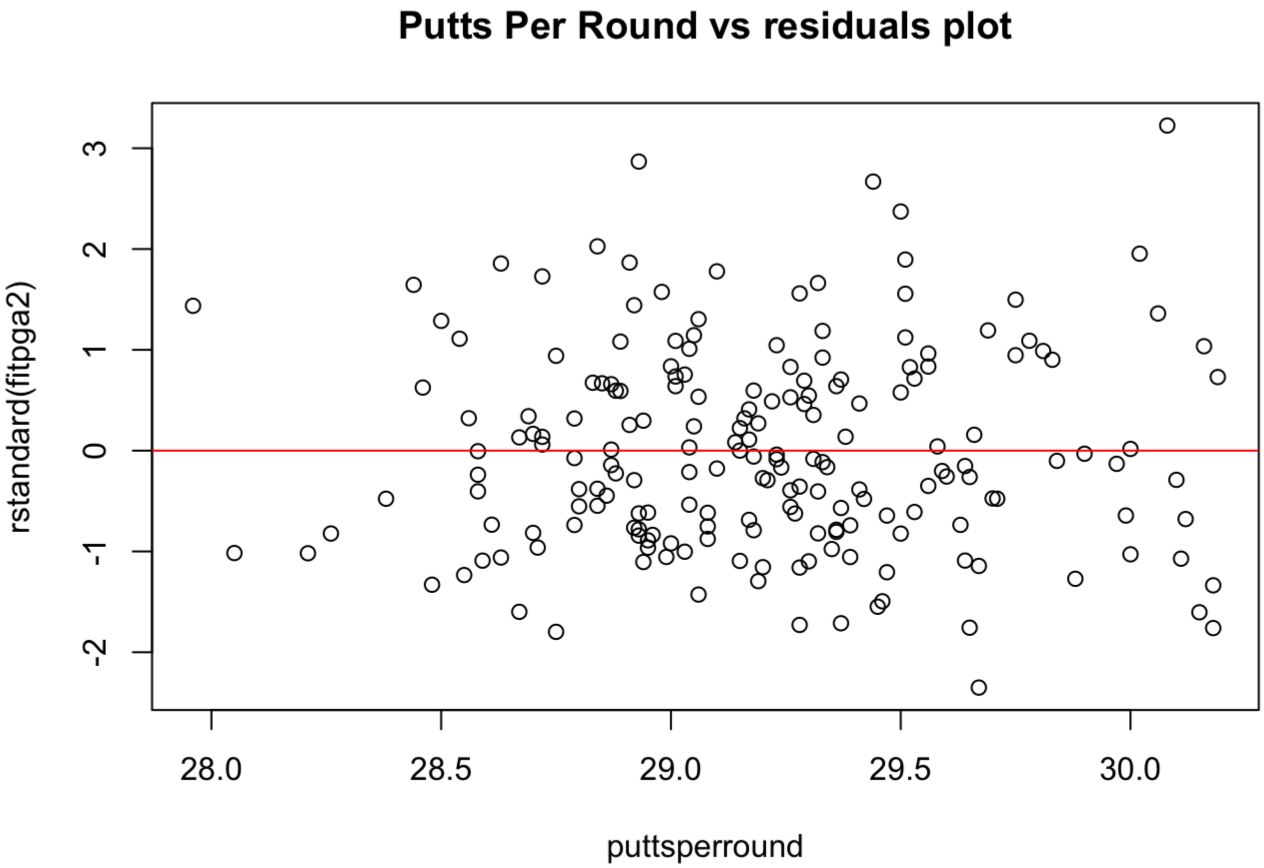

```
plot(birdieconversion, rstandard(fitpga2), main="Birdie Conversion vs residuals plot")
abline(a=0, b=0,col='red')
```



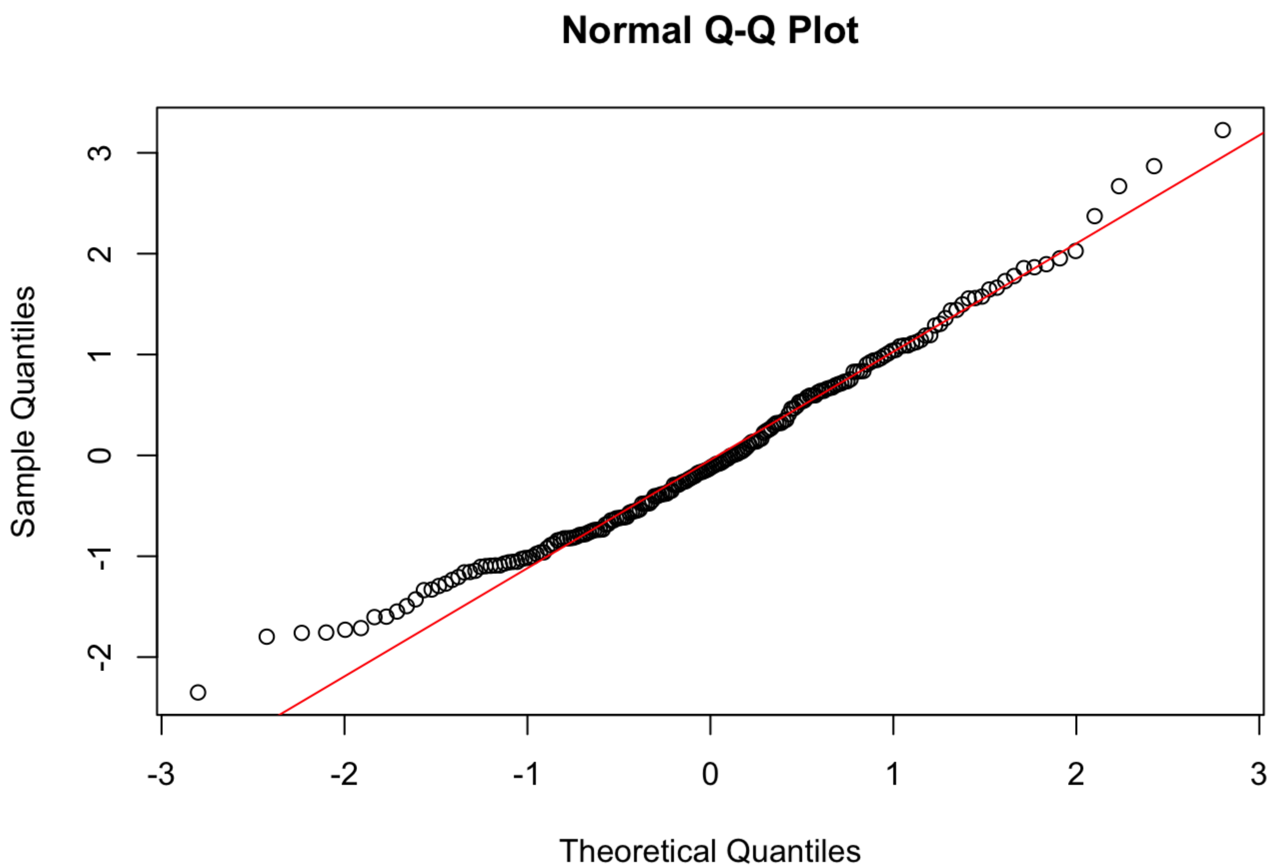

```
plot(puttingavg, rstandard(fitpga2), main="Putting Average vs residuals plot")
abline(a=0, b=0,col='red')
```




```
plot(puttsperround, rstandard(fitpga2), main="Putts Per Round vs residuals plot")
abline(a=0, b=0,col='red')
```



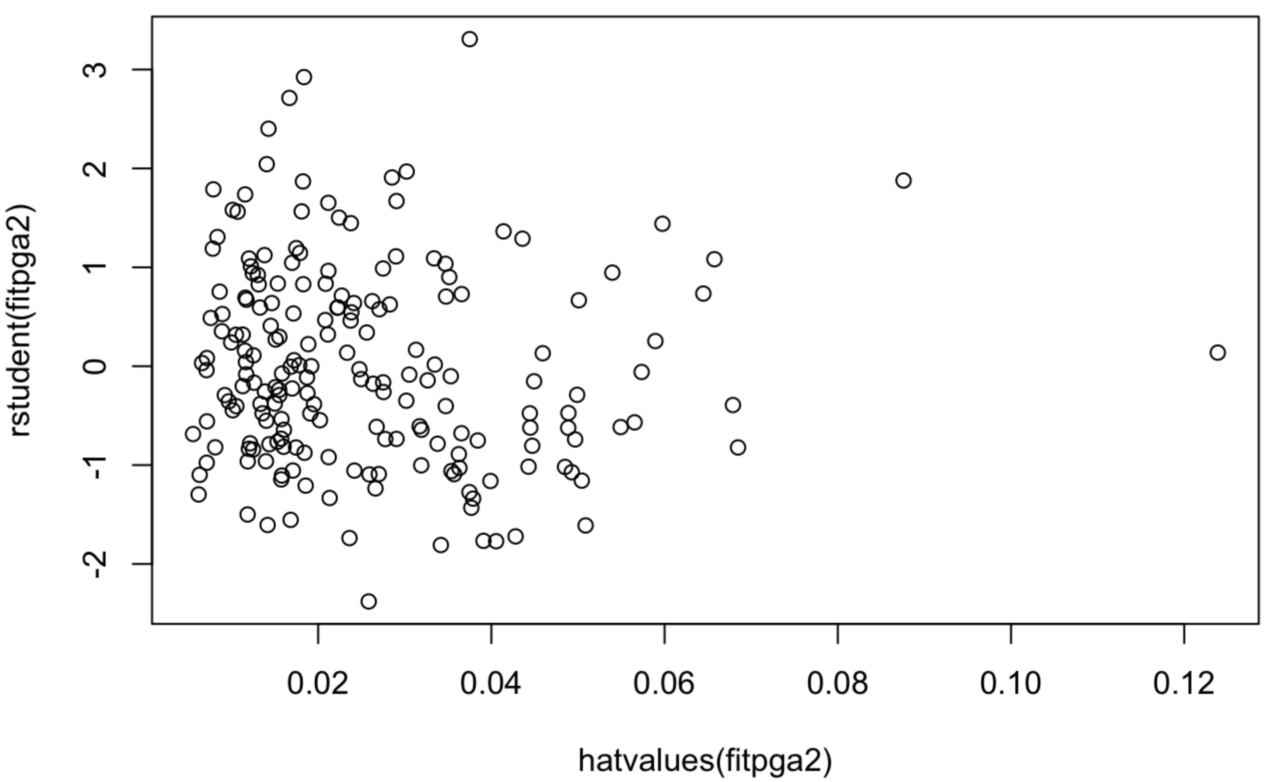

```
#normal probability plot of residuals
qqnorm(rstandard(fitpga2))
qqline(rstandard(fitpga2), col = 2)
```



Answer: Based on the residual plots, I do believe the model assumptions are met by the data. * Standardized residuals vs predicted: The first plot, predicted v. residuals, does appear to show a random scatter, so linearity seems to be satisfied. Constant variance appears to be confirmed. * Standardized residuals vs x-variables: The 4 variable plots also appear to be random, confirming linearity. * Normal plot of residuals: For the most part, the points are close to the line indicating normal distribution of errors. * Outliers: There may outliers, more analysis is needed.

*** Analyze if there are any outliers and influential points. If there are points in the dataset that need to be investigated, give one or more reason to support each point chosen. [1 pt R code, 1 pt answer = 2 pts]**

```
# plot of deleted studentized residuals vs hat values
plot(rstudent(fitpga2)~hatvalues(fitpga2))
```




```
rstudent(fitpga2)
```

```
##            1            2            3            4            5 
##  1.4407327802  1.5655155521 -1.1570812837 -1.4994128764 -0.8423254100 
##            6            7            8            9           10 
##  0.8335845487  1.3636760678  0.6247970772  2.4015472546 -0.6859203295 
##           11           12           13           14           15 
## -0.2119108818  0.3532095630  0.6386374014 -0.1671857659  0.3193145320 
##           16           17           18           19           20 
## -1.2077444583 -0.8756502414  1.2896299270 -0.2013668094 -0.8351956555 
##           21           22           23           24           25 
##  0.2698609737  0.5450094773  0.4614649110 -0.2920992021 -1.0160773924 
##           26           27           28           29           30 
## -1.0910661153 -0.8201056641 -0.3491615296  0.6928500413  2.0434201287 
##           31           32           33           34           35 
## -0.0306738401  1.5620698756  1.5023704840 -0.3840836521 -1.1054681072 
##           36           37           38           39           40 
##  0.9011411924  0.4879309256 -0.7335439507 -1.7700760780  1.8779395532 
##           41           42           43           44           45 
##  0.4666738014 -0.9625771554  0.5278731481 -0.0397290065 -1.0182609554 
##           46           47           48           49           50 
## -0.3773112865 -2.3790789898 -0.1429923367  1.0821331306 -1.2970889193 
##           51           52           53           54           55 
##  1.0892204959  0.8356182903  0.6725987777 -0.4031918369 -0.2921284134 
##           56           57           58           59           60 
## -0.6431078973 -0.5513214963  1.7374075667 -0.5346997610  1.9685804103 
##           61           62           63           64           65 
##  0.9226110329  0.3177903026  2.9238249758 -1.0607597761 -0.6133764640 
##           66           67           68           69           70 
## -0.1525432921 -1.2358882061  0.0416273110 -0.4039320821 -0.3926507740 
##           71           72           73           74           75 
## -0.2610162630 -0.3555932329  1.3064334548  0.9462785375  1.1938775557 
##           76           77           78           79           80 
##  1.0458273066 -0.5458613103 -0.7401636020  0.5947453751  0.5762618481 
##           81           82           83           84           85 
## -1.5539765589 -1.1448697632 -1.0721820716  1.1887982699 -0.7886462276 
##           86           87           88           89           90 
##  0.2405190661  0.1095906467 -0.9762275011 -0.0056683084  0.6669904667 
##           91           92           93           94           95 
## -1.0024880239  0.9882581989 -0.8150030600  0.0167419500 -0.0818791383 
##           96           97           98           99          100 
## -0.5679413184 -0.8047386038 -0.7640952856  0.2230985535 -0.1012281210 
##          101          102          103          104          105 
## -1.7660828349 -1.4313992332 -0.6427270467  1.7887721477  0.6579227880 
##          106          107          108          109          110 
## -0.9196716426  1.1235978500  0.7290879791 -0.2722004951 -0.8220196447 
##          111          112          113          114          115 
## -0.0736333507 -1.0987798634 -0.0582535980 -0.0846873853 -1.8083743878 
##          116          117          118          119          120 
##  0.0002950588  1.1107006452 -1.6056728807 -0.6228605851  0.7046648769 
##          121          122          123          124          125 
##  0.1659570594 -1.7380409463  1.0357184358 -1.0557874657 -1.6102973158 
##          126          127          128          129          130 
## -1.3327213398 -1.0285451185 -1.7211628551  0.5333841647  0.3208912760 
##          131          132          133          134          135
```

```
## -0.6777984956 -0.6156273510 -0.4456093096 -0.1634245945  1.0100604599
##           136           137           138           139           140
##  0.0322979457 -1.0550819700 -0.4763218634  0.4081117500 -0.4761871626
##           141           142           143           144           145
##  1.9085816391  0.2544592780  1.5808117510 -0.2558381508  0.2965444184
##           146           147           148           149           150
## -0.1772874582 -0.7797230046 -0.5577070519 -1.1608093269  1.6520862126
##           151           152           153           154           155
## -1.3390103644 -0.7355693003  1.4465911095 -0.7841907111  0.3405562010
##           156           157           158           159           160
## -0.3810047150  0.0104373742 -0.6074907611  0.0603194106  0.9637540442
##           161           162           163           164           165
##  1.0905339180 -0.9613863158  0.7135676196 -0.2384012340 -1.0945706692
##           166           167           168           169           170
##  1.1455014390 -0.8895043859 -1.2735582554 -0.2253011399 -0.1135038710
##           171           172           173           174           175
##  0.0826103527 -0.8214993121  0.9404797041  1.8688686301 -0.7367779830
##           176           177           178           179           180
##  0.1567291189 -0.2892025773  0.1374122907  0.7536712119  2.7130288291
##           181           182           183           184           185
##  0.6385877315 -0.4782361821 -1.0916735320  0.7345963697  3.3081097251
##           186           187           188           189           190
##  0.5932212096 -0.7512796320  0.8294596026  0.8275049089 -0.6213634666
##           191           192           193           194           195
##  0.1311040598  0.5912159683 -0.4731819302 -0.1303873005  0.1363268402
##           196
##  1.6705542137
```

Answer: There is one point that is outside of the (-3,3) bound for studentized residuals. This Point 185 also has the highest dffit, suggesting it has some influence on the model, as taking it out changes the predicted value by this level. All other points seem to be within reasonable bounds.

```
#run model diagnostics....
# compute influential points statistics
influence.measures(fitpga2)
```

```
## Influence measures of
##   lm(formula = ln_Prize ~ gir + birdieconversion + puttingavg +      puttsperround) :
##
##        dfb.1_  dfb.gir  dfb.brdc  dfb.pttn  dfb.ptts     dffit cov.r
## 1    6.13e-02 -6.47e-02  7.47e-02  1.01e-01 -1.69e-01  3.63e-01 1.034
## 2   -1.55e-02  1.30e-01  2.85e-02  2.58e-03 -2.13e-02  2.13e-01 0.981
## 3    4.15e-02  2.30e-01 -6.77e-02  4.01e-02 -1.24e-01 -2.67e-01 1.044
## 4   -8.03e-03 -1.93e-02  1.95e-02  6.47e-02 -7.13e-02 -1.64e-01 0.980
## 5   -6.73e-02  3.70e-02  4.65e-02  6.14e-02 -3.38e-02 -9.47e-02 1.020
## 6    1.64e-02  2.94e-02 -2.85e-02 -5.37e-02  4.97e-02  1.22e-01 1.030
## 7    5.65e-02 -1.96e-03 -1.76e-01 -1.30e-01  1.36e-01  2.83e-01 1.020
## 8    4.31e-02 -4.66e-02  9.76e-03 -3.54e-02  1.65e-02  1.07e-01 1.046
## 9    8.81e-02 -5.52e-02 -1.89e-01 -1.05e-01  9.30e-02  2.89e-01 0.897
## 10  -4.68e-03 -1.28e-02  4.27e-03 -2.37e-03  8.61e-03 -5.13e-02 1.020
## 11   1.27e-02 -8.04e-03 -1.42e-02 -2.03e-02  1.78e-02 -2.62e-02 1.041
## 12   3.50e-03  1.51e-02 -1.14e-02  3.40e-03 -8.58e-03  3.34e-02 1.032
## 13   8.59e-03  3.80e-02 -3.32e-02  1.42e-02 -2.78e-02  7.79e-02 1.031
## 14   8.87e-03 -3.00e-03 -1.23e-02 -4.13e-03 -6.84e-05 -1.89e-02 1.039
## 15   1.34e-02  1.67e-02 -1.65e-02 -3.23e-03 -8.37e-03  3.41e-02 1.036
## 16  -1.72e-02 -1.05e-01  5.41e-02  7.99e-03  1.93e-02 -1.66e-01 1.007
## 17  -9.34e-02  4.40e-02  9.39e-02  8.53e-02 -4.92e-02 -1.20e-01 1.025
## 18   1.08e-01 -1.52e-01 -6.22e-02 -2.32e-02 -1.88e-02  2.75e-01 1.028
## 19  -5.45e-04 -8.57e-04  6.45e-03  7.59e-03 -9.55e-03 -2.15e-02 1.037
## 20  -2.80e-02 -3.77e-02  2.88e-02 -2.12e-02  5.27e-02 -9.21e-02 1.020
## 21  -4.86e-03  2.72e-02  3.29e-03  1.41e-02 -1.93e-02  3.34e-02 1.040
## 22  -4.04e-02  1.94e-02  5.65e-02  1.40e-02  4.78e-03  8.52e-02 1.043
## 23   1.14e-03  2.11e-02 -2.38e-02  2.77e-02 -3.58e-02  7.20e-02 1.046
## 24   8.90e-03 -1.25e-02 -5.31e-03 -1.71e-02  1.67e-02 -2.82e-02 1.034
## 25  -1.34e-01 -4.22e-02  4.78e-02  1.49e-02  9.49e-02 -2.19e-01 1.045
## 26  -2.47e-02 -5.96e-02  5.03e-02  7.35e-02 -6.36e-02 -1.82e-01 1.023
## 27   1.30e-02  1.86e-02 -1.86e-02  1.65e-02 -3.27e-02 -7.42e-02 1.017
## 28   5.40e-02 -1.43e-02 -5.15e-02 -3.87e-02  1.25e-02 -6.16e-02 1.055
## 29  -4.33e-02 -7.12e-03  5.21e-02  2.04e-02  5.65e-03  7.51e-02 1.026
## 30  -3.08e-02 -3.99e-02  8.27e-02  1.03e-01 -1.01e-01  2.44e-01 0.934
## 31   2.84e-03  1.88e-03 -1.44e-03 -2.87e-04 -2.22e-03 -4.89e-03 1.053
## 32  -5.42e-02  8.65e-02  5.63e-03  4.94e-02 -3.41e-02  1.62e-01 0.974
## 33  -1.29e-01  5.25e-02  2.97e-02  1.24e-01 -5.86e-02  2.27e-01 0.990
## 34   1.41e-02 -4.20e-02 -7.99e-05 -2.80e-02  3.13e-02 -5.42e-02 1.043
## 35   8.72e-03 -3.83e-02 -5.99e-02 -4.75e-03  1.43e-02 -1.40e-01 1.010
## 36  -8.78e-02 -6.89e-02  7.85e-02 -2.30e-02  1.05e-01  1.72e-01 1.042
## 37   1.35e-02  5.54e-03 -1.18e-02 -1.67e-02  9.86e-03  4.27e-02 1.028
## 38  -4.94e-02  2.49e-02  1.51e-02  1.76e-02  1.16e-02 -9.27e-02 1.028
## 39   1.75e-02  4.00e-02  1.05e-01  1.70e-01 -2.45e-01 -3.64e-01 0.986
## 40   5.59e-02  2.57e-01 -1.75e-01  2.83e-01 -4.21e-01  5.82e-01 1.026
## 41   4.07e-03 -3.41e-02  3.07e-03 -4.09e-02  5.35e-02  6.80e-02 1.042
## 42   2.40e-02  3.38e-02 -5.26e-02 -2.79e-02  1.36e-02 -1.05e-01 1.014
## 43   6.11e-03 -3.19e-02 -4.55e-03 -2.10e-02  2.82e-02  5.01e-02 1.028
## 44   7.33e-05 -1.66e-03  3.06e-04 -7.89e-04  1.20e-03 -3.36e-03 1.034
## 45  -6.79e-02  3.24e-02 -5.95e-02  5.05e-02 -4.90e-03 -2.30e-01 1.050
## 46  -3.13e-02  2.41e-02  2.17e-02  2.38e-02 -1.15e-02 -4.66e-02 1.038
## 47  -8.28e-02 -1.55e-01  1.92e-01  1.24e-01 -7.67e-02 -3.87e-01 0.910
## 48   3.42e-05  2.06e-02 -7.07e-03  4.44e-03 -8.98e-03 -2.63e-02 1.061
## 49   1.17e-01 -1.90e-01 -2.82e-02 -2.19e-01  2.15e-01  2.87e-01 1.066
```

```
## 50  -3.12e-02 -1.78e-02  3.39e-02  2.10e-02 -1.11e-03 -1.02e-01 0.988
## 51  -1.17e-02  2.21e-02  5.25e-02  2.83e-03 -5.06e-03  1.20e-01 1.007
## 52   2.27e-02  4.89e-03  1.26e-02 -3.82e-02  2.54e-02  1.04e-01 1.024
## 53   2.78e-02  3.19e-02 -1.41e-02  9.21e-03 -3.88e-02  7.33e-02 1.026
## 54   2.90e-03  1.40e-03 -1.27e-02 -4.07e-02  4.83e-02 -7.65e-02 1.059
## 55   9.37e-03 -1.15e-02 -1.20e-02 -2.56e-02  2.74e-02 -3.67e-02 1.040
## 56   5.81e-02 -3.68e-02 -3.97e-02 -6.15e-02  3.98e-02 -8.21e-02 1.032
## 57  -2.80e-02 -7.97e-03  1.81e-03  1.98e-02  2.40e-04 -6.57e-02 1.033
## 58   4.00e-02  4.43e-03  3.40e-02  2.48e-02 -6.79e-02  1.88e-01 0.960
## 59  -1.28e-02 -4.66e-02  4.96e-03 -3.32e-03  2.37e-02 -6.77e-02 1.035
## 60  -1.02e-01 -1.73e-01 -1.45e-02 -8.85e-02  2.32e-01  3.47e-01 0.957
## 61   5.19e-02  1.59e-02 -7.63e-02 -4.00e-02  1.32e-02  1.06e-01 1.017
## 62   1.94e-02 -2.74e-03 -9.45e-03 -8.41e-03 -3.74e-03  3.27e-02 1.035
## 63  -1.52e-01 -1.13e-01  2.77e-01  1.24e-01 -3.81e-02  4.00e-01 0.839
## 64  -9.17e-02  1.49e-01  4.70e-02  6.70e-02 -4.46e-02 -2.03e-01 1.033
## 65  -8.15e-02  4.64e-02  7.97e-02  6.67e-02 -3.49e-02 -1.02e-01 1.044
## 66   2.60e-02 -2.14e-02 -1.99e-02 -2.74e-02  1.90e-02 -3.31e-02 1.074
## 67   3.37e-02 -3.85e-02 -1.24e-01 -8.13e-02  9.55e-02 -2.04e-01 1.013
## 68  -2.54e-04  1.50e-03 -1.15e-03 -5.60e-04  7.42e-04  4.52e-03 1.039
## 69   1.03e-02 -3.88e-03  8.70e-04 -1.85e-02  1.46e-02 -4.17e-02 1.033
## 70  -1.14e-02  5.72e-02  3.31e-02 -2.35e-03 -4.65e-03 -1.06e-01 1.097
## 71   1.37e-02  1.07e-02 -1.50e-02  1.31e-02 -2.77e-02 -4.39e-02 1.054
## 72   2.69e-03  1.72e-02  2.06e-03 -7.62e-04 -5.42e-03 -3.52e-02 1.033
## 73   4.61e-02  4.93e-02 -4.04e-02 -7.74e-03 -3.32e-02  1.20e-01 0.990
## 74  -1.88e-01 -3.96e-02  1.86e-01  8.77e-02  3.22e-02  2.26e-01 1.060
## 75   1.39e-02  4.42e-03 -5.24e-02 -7.60e-02  8.77e-02  1.59e-01 1.007
## 76   9.39e-03 -2.95e-02 -4.15e-02  2.46e-02 -2.53e-02  1.38e-01 1.015
## 77   1.52e-02  1.26e-02 -4.89e-02 -6.50e-03 -1.06e-03 -7.84e-02 1.040
## 78   1.22e-01  3.45e-02 -1.10e-01 -1.13e-01  4.33e-02 -1.69e-01 1.065
## 79  -3.44e-02  5.78e-02  2.16e-02  7.05e-02 -7.24e-02  8.98e-02 1.040
## 80  -5.47e-02 -5.36e-02  5.43e-02  1.30e-02  3.38e-02  9.61e-02 1.046
## 81   1.10e-01  5.67e-02 -6.61e-02 -9.05e-02  1.56e-02 -2.03e-01 0.980
## 82   3.68e-02  2.24e-02 -1.71e-02  4.50e-02 -8.81e-02 -1.45e-01 1.008
## 83   2.56e-02 -1.33e-01  7.24e-02  4.94e-03 -8.72e-03 -2.44e-01 1.048
## 84   2.71e-02 -1.30e-03 -5.35e-02 -2.10e-02  1.12e-02  1.06e-01 0.997
## 85   4.18e-02 -4.84e-02 -3.19e-02 -7.33e-02  6.90e-02 -9.52e-02 1.025
## 86   1.44e-02 -5.36e-03 -9.79e-03 -1.53e-02  9.36e-03  2.41e-02 1.035
## 87  -3.29e-05  9.14e-03 -1.05e-03  4.34e-03 -7.03e-03  1.23e-02 1.039
## 88  -1.06e-02 -8.98e-03  1.90e-02  2.54e-02 -2.32e-02 -8.27e-02 1.008
## 89  -3.95e-04 -7.13e-05  7.69e-05  1.33e-04  1.65e-04 -7.42e-04 1.044
## 90   5.57e-02  1.09e-01 -3.57e-02 -1.53e-02 -4.57e-02  1.53e-01 1.068
## 91  -1.09e-01  1.43e-01  8.70e-02  1.42e-01 -1.25e-01 -1.82e-01 1.033
## 92   3.18e-03  2.12e-02 -3.56e-02 -7.43e-02  8.84e-02  1.66e-01 1.029
## 93  -1.19e-03 -4.47e-02 -3.54e-02 -4.01e-02  6.38e-02 -1.04e-01 1.025
## 94   2.83e-04 -4.43e-04 -1.25e-03 -1.74e-03  2.16e-03  3.12e-03 1.062
## 95  -2.68e-03 -3.16e-03  5.02e-03  5.52e-04  1.51e-03 -8.90e-03 1.039
## 96   7.01e-02  8.88e-02 -8.54e-02 -1.51e-02 -4.73e-02 -1.39e-01 1.079
## 97  -4.06e-02  1.25e-01  7.30e-02  4.88e-02 -6.28e-02 -1.74e-01 1.057
## 98   3.31e-02  1.80e-02 -5.45e-02 -4.56e-02  3.13e-02 -9.53e-02 1.027
## 99   1.25e-02 -9.59e-03 -5.72e-03 -2.24e-02  1.98e-02  3.09e-02 1.045
## 100 -1.27e-03 -9.25e-03  6.96e-03  4.42e-03 -3.33e-03 -1.94e-02 1.064
## 101  1.02e-01  3.56e-02  3.90e-02 -1.42e-01  7.46e-02 -3.56e-01 0.985
## 102  7.74e-02 -1.59e-01 -1.46e-01 -6.19e-02  6.46e-02 -2.83e-01 1.011
## 103  6.08e-02 -5.29e-02 -5.17e-03 -5.19e-02  2.48e-02 -1.17e-01 1.049
```

```
## 104  8.98e-02 -1.20e-03 -7.74e-02 -7.46e-02  2.81e-02  1.60e-01 0.952
## 105  7.08e-02  2.08e-03 -7.79e-02 -1.95e-02 -2.44e-02  1.08e-01 1.042
## 106  3.45e-02 -8.32e-02 -6.73e-02 -4.66e-02  5.48e-02 -1.35e-01 1.026
## 107 -1.08e-03  7.41e-02 -3.93e-02  5.29e-03 -1.57e-02  1.33e-01 1.007
## 108 -1.54e-02 -1.55e-02 -4.06e-02 -5.42e-02  8.77e-02  1.42e-01 1.051
## 109 -2.11e-02  1.65e-02  1.68e-02  3.16e-02 -2.70e-02 -3.76e-02 1.044
## 110  7.85e-03 -7.37e-02  2.68e-02 -2.45e-02  3.51e-02 -1.10e-01 1.026
## 111 -6.89e-03 -1.79e-03  5.11e-03  3.23e-03  1.41e-03 -9.34e-03 1.043
## 112  4.31e-03 -3.17e-02  6.76e-03 -6.55e-03  9.84e-03 -8.76e-02 1.001
## 113 -7.67e-03 -1.55e-03  1.15e-02  2.06e-03  2.58e-03 -1.44e-02 1.089
## 114 -2.95e-03  1.20e-02  3.99e-04  9.68e-03 -1.20e-02 -1.50e-02 1.059
## 115 -7.07e-02  2.58e-01  5.88e-03  5.22e-02 -6.22e-02 -3.40e-01 0.976
## 116 -1.67e-05  1.34e-05  2.68e-05  9.76e-06 -4.42e-06  4.13e-05 1.047
## 117  4.98e-02  1.24e-01 -3.56e-03  5.91e-02 -1.39e-01  1.92e-01 1.024
## 118 -6.41e-02 -5.66e-02 -1.08e-02 -1.24e-02  8.08e-02 -1.92e-01 0.973
## 119 -8.80e-02  3.74e-02  1.22e-01  6.10e-02 -2.63e-02 -1.41e-01 1.068
## 120 -7.56e-02 -7.43e-02  8.56e-02  2.70e-02  3.43e-02  1.34e-01 1.050
## 121  1.16e-02 -1.57e-02 -8.69e-03 -2.56e-03 -1.84e-03  2.98e-02 1.059
## 122 -1.31e-01  3.35e-03  1.10e-01  1.95e-01 -1.46e-01 -2.70e-01 0.972
## 123 -1.40e-01 -2.00e-02  7.16e-02  3.88e-02  6.30e-02  1.96e-01 1.034
## 124  1.26e-01 -8.91e-02 -1.06e-01 -1.40e-01  1.00e-01 -1.66e-01 1.022
## 125  3.30e-01 -5.84e-02 -2.03e-01 -2.29e-01  4.63e-02 -3.73e-01 1.011
## 126 -3.42e-02  1.90e-02 -6.15e-02 -2.09e-02  5.72e-02 -1.97e-01 1.001
## 127  3.60e-02  1.24e-01  2.89e-02  6.27e-02 -1.37e-01 -2.00e-01 1.036
## 128  2.25e-01 -7.47e-02 -1.51e-01 -3.04e-01  2.20e-01 -3.64e-01 0.993
## 129  3.36e-02 -8.92e-03 -1.58e-02 -4.68e-02  3.36e-02  7.04e-02 1.037
## 130 -3.17e-03  5.92e-03  2.16e-02  2.08e-02 -2.67e-02  4.71e-02 1.046
## 131  8.60e-02  4.04e-02 -3.96e-02 -2.19e-02 -4.79e-02 -1.32e-01 1.053
## 132  1.13e-01 -5.89e-02 -1.28e-01 -1.28e-01  9.21e-02 -1.48e-01 1.075
## 133 -2.92e-02  1.18e-03  1.98e-02  1.34e-02  4.72e-03 -4.51e-02 1.032
## 134  2.84e-03 -1.64e-02  5.62e-03 -1.41e-02  1.75e-02 -2.75e-02 1.055
## 135  1.61e-02 -3.08e-02  2.22e-02 -4.35e-02  4.39e-02  1.12e-01 1.012
## 136 -1.75e-04 -1.39e-04  8.52e-04  2.89e-04 -2.77e-04  2.63e-03 1.033
## 137 -3.77e-02 -3.27e-02  5.36e-02 -3.60e-02  7.43e-02 -1.39e-01 1.014
## 138 -2.72e-02 -3.56e-02  1.52e-02 -4.27e-02  7.98e-02 -1.03e-01 1.068
## 139  3.78e-03  2.22e-02 -1.43e-02  1.87e-02 -2.85e-02  4.95e-02 1.037
## 140  2.33e-02  3.00e-04  1.19e-03 -1.09e-02 -6.31e-03 -5.59e-02 1.035
## 141 -1.02e-01 -5.61e-03 -1.63e-02  1.56e-01 -1.02e-01  3.27e-01 0.961
## 142 -1.56e-02  1.01e-02  3.90e-02  1.07e-03  4.26e-03  6.37e-02 1.089
## 143 -4.09e-02  6.96e-02  7.57e-02  8.15e-02 -8.99e-02  1.60e-01 0.972
## 144  6.53e-03  1.08e-02 -3.12e-03  1.08e-02 -2.06e-02 -3.03e-02 1.039
## 145  2.60e-02  1.05e-02 -2.35e-02 -1.30e-02 -4.66e-03  3.72e-02 1.040
## 146 -1.33e-02 -5.80e-04  7.01e-03  1.90e-02 -1.31e-02 -2.92e-02 1.054
## 147 -2.47e-02 -9.20e-03  2.44e-02 -1.91e-02  4.20e-02 -8.64e-02 1.023
## 148 -1.06e-02  2.55e-03  2.00e-02  4.32e-03  3.16e-06 -4.74e-02 1.026
## 149  1.97e-01 -2.90e-02 -2.15e-01 -1.72e-01  8.13e-02 -2.37e-01 1.032
## 150  1.37e-01 -3.75e-02 -3.39e-02 -4.28e-02 -4.77e-02  2.43e-01 0.977
## 151  1.50e-01  8.46e-02 -4.41e-02 -2.59e-02 -1.05e-01 -2.66e-01 1.018
## 152 -6.23e-02  4.17e-02  9.56e-02  9.11e-02 -8.16e-02 -1.27e-01 1.042
## 153 -1.14e-01  9.69e-02  1.42e-01  1.87e-01 -1.72e-01  2.26e-01 0.996
## 154  2.12e-02  1.03e-01 -1.51e-03 -4.19e-03 -3.47e-02 -1.47e-01 1.045
## 155 -1.60e-02  1.72e-02  2.76e-02  4.00e-02 -4.28e-02  5.52e-02 1.050
## 156  8.04e-03 -8.36e-03 -2.34e-02 -2.00e-02  2.24e-02 -4.43e-02 1.037
## 157  1.13e-03  6.82e-05 -1.02e-03 -6.19e-04 -5.01e-05  1.41e-03 1.045
```

```
## 158 -4.40e-02  4.73e-02  4.68e-02  9.28e-02 -9.34e-02 -1.10e-01 1.050
## 159  3.37e-03  4.19e-03 -1.39e-03  4.64e-04 -4.03e-03  7.98e-03 1.044
## 160 -1.11e-01 -2.79e-02  1.00e-01  5.91e-02  1.28e-02  1.42e-01 1.024
## 161 -1.75e-01  2.85e-02  1.53e-01  9.39e-02  3.44e-03  2.03e-01 1.029
## 162 -7.44e-02  2.23e-02  3.11e-02  4.85e-02 -7.01e-03 -1.14e-01 1.016
## 163  4.91e-02 -1.31e-02 -6.42e-02 -8.05e-02  6.92e-02  1.09e-01 1.037
## 164 -1.30e-02  6.02e-04 -2.88e-04  2.39e-03  7.59e-03 -2.99e-02 1.041
## 165 -1.26e-01 -4.27e-02  1.42e-01  9.00e-02 -1.57e-02 -1.79e-01 1.021
## 166  8.72e-02 -1.75e-02 -5.23e-02 -1.09e-01  7.24e-02  1.55e-01 1.010
## 167 -1.41e-01 -3.01e-02  1.36e-01  1.03e-01 -2.19e-02 -1.72e-01 1.043
## 168  1.47e-02 -1.37e-01  9.85e-02 -2.97e-02  3.97e-02 -2.51e-01 1.022
## 169 -9.51e-03 -1.82e-02  8.21e-03 -6.88e-03  1.89e-02 -2.96e-02 1.043
## 170  2.06e-03  2.61e-03  2.35e-03 -5.08e-03  3.56e-03 -1.57e-02 1.046
## 171  7.80e-04  2.67e-03 -1.29e-03  1.59e-03 -2.95e-03  7.01e-03 1.034
## 172 -7.88e-02 -1.62e-01  3.12e-02 -4.51e-02  1.50e-01 -2.23e-01 1.083
## 173  1.22e-02  2.78e-02  2.99e-02  2.02e-02 -4.42e-02  1.06e-01 1.016
## 174  1.21e-01 -7.44e-02  1.76e-04 -9.61e-02  3.44e-02  2.55e-01 0.955
## 175  3.41e-02  5.58e-02 -7.51e-02 -2.89e-02  4.90e-03 -1.24e-01 1.041
## 176 -8.00e-03  1.05e-03  3.54e-03  7.13e-04  5.09e-03  1.69e-02 1.038
## 177  3.13e-02 -1.10e-02  2.42e-03 -3.01e-02  1.28e-02 -6.63e-02 1.078
## 178 -2.77e-02  3.38e-02  3.41e-02  2.26e-02 -1.63e-02  5.17e-02 1.171
## 179  2.56e-03  1.36e-02  1.85e-02 -2.74e-03 -3.83e-03  7.03e-02 1.020
## 180 -2.07e-02 -2.29e-01 -5.48e-02 -4.69e-02  1.32e-01  3.53e-01 0.863
## 181 -1.65e-02 -1.59e-02  9.65e-03  4.96e-02 -4.44e-02  1.00e-01 1.041
## 182 -8.35e-03 -3.75e-03  3.35e-02 -7.06e-03  1.16e-02 -6.67e-02 1.040
## 183 -1.82e-01  9.77e-02  1.29e-01  1.50e-01 -7.17e-02 -2.10e-01 1.032
## 184  3.54e-02 -1.68e-01 -3.04e-02 -5.00e-02  7.21e-02  1.93e-01 1.082
## 185  6.40e-03 -1.75e-01 -3.19e-01 -1.92e-01  3.10e-01  6.53e-01 0.806
## 186 -2.50e-03 -9.85e-03  3.19e-02 -1.79e-03  2.34e-03  6.88e-02 1.031
## 187 -9.02e-02  9.93e-02  6.47e-02  1.33e-01 -1.18e-01 -1.50e-01 1.052
## 188 -1.65e-02  8.38e-02  1.93e-02  2.24e-02 -3.39e-02  1.13e-01 1.027
## 189 -6.32e-02  3.75e-02  3.66e-02  6.06e-02 -3.42e-02  9.55e-02 1.022
## 190 -7.56e-02  1.08e-01  4.80e-02  1.07e-01 -9.76e-02 -1.34e-01 1.064
## 191  6.34e-03 -1.17e-02  6.77e-03 -1.33e-02  1.26e-02  2.88e-02 1.076
## 192 -8.28e-03  5.59e-02  3.34e-02  2.32e-02 -3.76e-02  8.91e-02 1.040
## 193  1.46e-02  7.71e-02 -1.73e-02  5.25e-02 -9.04e-02 -1.07e-01 1.073
## 194  1.41e-02 -1.27e-03 -8.12e-03 -3.59e-03 -5.98e-03 -2.09e-02 1.052
## 195 -6.75e-03  2.90e-03  1.18e-02  1.41e-02 -1.38e-02  2.11e-02 1.051
## 196  2.06e-01 -1.31e-02 -2.44e-01 -2.06e-01  1.17e-01  2.89e-01 0.983
##       cook.d     hat inf
## 1   2.62e-02 0.05975
## 2   8.97e-03 0.01810
## 3   1.42e-02 0.05044
## 4   5.36e-03 0.01185
## 5   1.80e-03 0.01248
## 6   2.97e-03 0.02087
## 7   1.60e-02 0.04140
## 8   2.28e-03 0.02827
## 9   1.63e-02 0.01426   *
## 10  5.27e-04 0.00555
## 11  1.38e-04 0.01507
## 12  2.25e-04 0.00889
## 13  1.22e-03 0.01465
## 14  7.18e-05 0.01261
```

```
## 15  2.34e-04 0.01128
## 16  5.50e-03 0.01856
## 17  2.88e-03 0.01842
## 18  1.51e-02 0.04359
## 19  9.31e-05 0.01129
## 20  1.70e-03 0.01200
## 21  2.24e-04 0.01509
## 22  1.46e-03 0.02383
## 23  1.04e-03 0.02375
## 24  1.60e-04 0.00922
## 25  9.57e-03 0.04429
## 26  6.60e-03 0.02700
## 27  1.10e-03 0.00811
## 28  7.62e-04 0.03018
## 29  1.13e-03 0.01160
## 30  1.17e-02 0.01406
## 31  4.80e-06 0.02476
## 32  5.24e-03 0.01070
## 33  1.03e-02 0.02241
## 34  5.90e-04 0.01952
## 35  3.92e-03 0.01581
## 36  5.92e-03 0.03514
## 37  3.66e-04 0.00759
## 38  1.72e-03 0.01573
## 39  2.62e-02 0.04055
## 40  6.68e-02 0.08759   *
## 41  9.30e-04 0.02082
## 42  2.23e-03 0.01187
## 43  5.05e-04 0.00894
## 44  2.27e-06 0.00711
## 45  1.06e-02 0.04850
## 46  4.35e-04 0.01500
## 47  2.93e-02 0.02584   *
## 48  1.39e-04 0.03264
## 49  1.65e-02 0.06575
## 50  2.09e-03 0.00620
## 51  2.88e-03 0.01201
## 52  2.18e-03 0.01534
## 53  1.08e-03 0.01172
## 54  1.17e-03 0.03473
## 55  2.70e-04 0.01551
## 56  1.35e-03 0.01605
## 57  8.67e-04 0.01401
## 58  6.99e-03 0.01157
## 59  9.19e-04 0.01576
## 60  2.38e-02 0.03022
## 61  2.25e-03 0.01303
## 62  2.15e-04 0.01050
## 63  3.08e-02 0.01836   *
## 64  8.25e-03 0.03539
## 65  2.07e-03 0.02675
## 66  2.20e-04 0.04493
## 67  8.33e-03 0.02661
## 68  4.11e-06 0.01166
```

```
## 69  3.49e-04 0.01055
## 70  2.26e-03 0.06789   *
## 71  3.88e-04 0.02754
## 72  2.48e-04 0.00968
## 73  2.87e-03 0.00837
## 74  1.02e-02 0.05397
## 75  5.06e-03 0.01747
## 76  3.78e-03 0.01700
## 77  1.23e-03 0.02021
## 78  5.74e-03 0.04967
## 79  1.62e-03 0.02229
## 80  1.85e-03 0.02707
## 81  8.20e-03 0.01682
## 82  4.18e-03 0.01574
## 83  1.19e-02 0.04928
## 84  2.23e-03 0.00783
## 85  1.82e-03 0.01437
## 86  1.17e-04 0.00996
## 87  3.06e-05 0.01253
## 88  1.37e-03 0.00713
## 89  1.11e-07 0.01683
## 90  4.71e-03 0.05010
## 91  6.62e-03 0.03190
## 92  5.52e-03 0.02749
## 93  2.17e-03 0.01601
## 94  1.95e-06 0.03347
## 95  1.59e-05 0.01169
## 96  3.88e-03 0.05654   *
## 97  6.07e-03 0.04472
## 98  1.82e-03 0.01530
## 99  1.92e-04 0.01886
## 100 7.54e-05 0.03531
## 101 2.51e-02 0.03909
## 102 1.60e-02 0.03769
## 103 2.73e-03 0.03192
## 104 5.04e-03 0.00790
## 105 2.34e-03 0.02625
## 106 3.67e-03 0.02121
## 107 3.53e-03 0.01380
## 108 4.05e-03 0.03657
## 109 2.85e-04 0.01876
## 110 2.40e-03 0.01744
## 111 1.75e-05 0.01583
## 112 1.53e-03 0.00632
## 113 4.15e-05 0.05735   *
## 114 4.54e-05 0.03052
## 115 2.29e-02 0.03417
## 116 3.43e-10 0.01920
## 117 7.36e-03 0.02899
## 118 7.35e-03 0.01417
## 119 4.00e-03 0.04887
## 120 3.59e-03 0.03477
## 121 1.79e-04 0.03130
## 122 1.45e-02 0.02362
```

```
## 123 7.71e-03 0.03470
## 124 5.52e-03 0.02419
## 125 2.76e-02 0.05088
## 126 7.71e-03 0.02133
## 127 7.97e-03 0.03631
## 128 2.62e-02 0.04280
## 129 9.96e-04 0.01714
## 130 4.46e-04 0.02112
## 131 3.50e-03 0.03655
## 132 4.42e-03 0.05493
## 133 4.09e-04 0.01014
## 134 1.52e-04 0.02751
## 135 2.53e-03 0.01223
## 136 1.39e-06 0.00660
## 137 3.87e-03 0.01708
## 138 2.12e-03 0.04444
## 139 4.93e-04 0.01452
## 140 6.27e-04 0.01357
## 141 2.11e-02 0.02853
## 142 8.15e-04 0.05894   *
## 143 5.08e-03 0.01014
## 144 1.85e-04 0.01385
## 145 2.79e-04 0.01552
## 146 1.71e-04 0.02635
## 147 1.50e-03 0.01214
## 148 4.50e-04 0.00716
## 149 1.12e-02 0.03989
## 150 1.17e-02 0.02117
## 151 1.41e-02 0.03789
## 152 3.24e-03 0.02904
## 153 1.01e-02 0.02377
## 154 4.31e-03 0.03379
## 155 6.13e-04 0.02561
## 156 3.94e-04 0.01334
## 157 3.97e-07 0.01782
## 158 2.43e-03 0.03171
## 159 1.28e-05 0.01721
## 160 4.02e-03 0.02118
## 161 8.21e-03 0.03338
## 162 2.62e-03 0.01397
## 163 2.37e-03 0.02273
## 164 1.80e-04 0.01550
## 165 6.37e-03 0.02591
## 166 4.77e-03 0.01790
## 167 5.96e-03 0.03624
## 168 1.26e-02 0.03747
## 169 1.76e-04 0.01700
## 170 4.94e-05 0.01871
## 171 9.89e-06 0.00715
## 172 9.94e-03 0.06848   *
## 173 2.23e-03 0.01245
## 174 1.28e-02 0.01825
## 175 3.11e-03 0.02775
## 176 5.77e-05 0.01155
```

```
## 177 8.83e-04 0.04990
## 178 5.37e-04 0.12388   *
## 179 9.91e-04 0.00863
## 180 2.42e-02 0.01668   *
## 181 2.02e-03 0.02413
## 182 8.94e-04 0.01910
## 183 8.81e-03 0.03569
## 184 7.46e-03 0.06448   *
## 185 8.11e-02 0.03750   *
## 186 9.51e-04 0.01328
## 187 4.52e-03 0.03843
## 188 2.57e-03 0.01828
## 189 1.83e-03 0.01313
## 190 3.61e-03 0.04449
## 191 1.66e-04 0.04593
## 192 1.59e-03 0.02221
## 193 2.31e-03 0.04889
## 194 8.75e-05 0.02497
## 195 8.93e-05 0.02335
## 196 1.65e-02 0.02905
```

```
# print out only observations that may be influential
summary(influence.measures(fitpga2))
```

```
## Potentially influential observations of
##   lm(formula = ln_Prize ~ gir + birdieconversion + puttingavg +      puttsperround) :
##
##     dfb.1_ dfb.gir dfb.brdc dfb.pttn dfb.ptts dffit   cov.r   cook.d
## 9    0.09  -0.06   -0.19    -0.10     0.09     0.29    0.90_*  0.02
## 40   0.06   0.26   -0.18     0.28    -0.42     0.58_*  1.03    0.07
## 47  -0.08  -0.16    0.19     0.12    -0.08    -0.39    0.91_*  0.03
## 63  -0.15  -0.11    0.28     0.12    -0.04     0.40    0.84_*  0.03
## 70  -0.01   0.06    0.03     0.00     0.00    -0.11    1.10_*  0.00
## 96   0.07   0.09   -0.09    -0.02    -0.05    -0.14    1.08_*  0.00
## 113 -0.01   0.00    0.01     0.00     0.00    -0.01    1.09_*  0.00
## 142 -0.02   0.01    0.04     0.00     0.00     0.06    1.09_*  0.00
## 172 -0.08  -0.16    0.03    -0.05     0.15    -0.22    1.08_*  0.01
## 178 -0.03   0.03    0.03     0.02    -0.02     0.05    1.17_*  0.00
## 180 -0.02  -0.23   -0.05    -0.05     0.13     0.35    0.86_*  0.02
## 184  0.04  -0.17   -0.03    -0.05     0.07     0.19    1.08_*  0.01
## 185  0.01  -0.17   -0.32    -0.19     0.31     0.65_*  0.81_*  0.08
##     hat
## 9    0.01
## 40   0.09_*
## 47   0.03
## 63   0.02
## 70   0.07
## 96   0.06
## 113  0.06
## 142  0.06
## 172  0.07
## 178  0.12_*
## 180  0.02
## 184  0.06
## 185  0.04
```

Problem 2 e)

Interpret the regression coefficients in the final model to answer the following question: How does an increase in 1% for GIR affect the average Prize money? [1 pt R code, 1 pt answer = 2 pts]

```
# regression coefficients
summary(fitpga2)
```

```
##
## Call:
## lm(formula = ln_Prize ~ gir + birdieconversion + puttingavg +
##     puttsperround)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.55608 -0.51122 -0.08109  0.45250  2.12227
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       8.02738    6.35383   1.263   0.2080
## gir               0.26791    0.02536  10.563  < 2e-16 ***
## birdieconversion  0.15360    0.03561   4.314 2.57e-05 ***
## puttingavg        8.81065    5.26991   1.672   0.0962 .
## puttsperround    -1.20702    0.26391  -4.574 8.61e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6707 on 191 degrees of freedom
## Multiple R-squared:  0.5414, Adjusted R-squared:  0.5318
## F-statistic: 56.37 on 4 and 191 DF,  p-value: < 2.2e-16
```

Answer:

From the table above, Beta1 = parameter estimate for gir = 0.26791, Beta2 = parameter estimate for birdieconversion = 0.26791, Beta3 = parameter estimate for puttingavg = 8.81065, Beta4 = parameter estimate for puttsperround = -1.20702, and Intercept = 8.02738

The fitted model is: log(PrizeMoney) = 8.02738 + 0.26791(gir) + 0.26791(birdieconversion) + 8.81065(puttingavg) - 1.20702(puttsperround) + e

So the GIR parameter coefficient indicates that if birdie conversion, putting average, and putts per round are fixed, a 1% change or change of 1 in GIR will increase log(PrizeMoney) by 0.268. Or Prize money by 1.3070 * 10000 = $13,070.

```
#turn back the log transform
exp(0.26791)
```

```
## [1] 1.307229
```

```
newf1 = data.frame(gir=c(67), puttingavg=c(1.77),birdieconversion=c(28), puttsperround=c
(29.16))
# compute average response value and confidence interval
predict(fitpga2, newf1, interval="confidence",level=0.95)
```

```
##        fit      lwr      upr
## 1 10.67594 10.50526 10.84661
```

```
exp(predict(fitpga2, newf1, interval="confidence",level=0.95))
```

```
##        fit      lwr      upr
## 1 43301.34 36507.18 51359.93
```

```
newf2 = data.frame(gir=c(68), puttingavg=c(1.77),birdieconversion=c(28), puttsperround=c
(29.16))
# compute average response value and confidence interval
predict(fitpga2, newf2, interval="confidence",level=0.95)
```

```
##        fit      lwr      upr
## 1 10.94385 10.75422 11.13348
```

```
exp(predict(fitpga2, newf2, interval="confidence",level=0.95))
```

```
##        fit      lwr      upr
## 1 56604.65 46827.03 68423.86
```

```
56604.65-43301.34
```

```
## [1] 13303.31
```

Answer: ^This confirms the value above for a 1% change in GIR - prize value increases by about $13000.

Problem 2 f)

Compute the prediction and 95% prediction interval for average prize money for a player that has a *GIR of 67%* driving accuracy of 64% *putting average of 1.77* Birdie Conversion of 28% *29.16 average putts per round. **[1 pt R code, 1 pt answer = 2 pts]**

```
new2 = data.frame(gir=c(67), puttingavg=c(1.77),birdieconversion=c(28), puttsperround=c(
29.16))
# compute average response value and confidence interval
predict(fitpga2, new2, interval="confidence",level=0.95)
```

```
##        fit      lwr      upr
## 1 10.67594 10.50526 10.84661
```

```
exp(predict(fitpga2, new2, interval="confidence",level=0.95))
```

```
##        fit      lwr      upr
## 1 43301.34 36507.18 51359.93
```

Answer: So the average prize money for those values would be 43044.94 dollars based on my model. The lower bound of 36507.18 and upper bound of 51359.93 for the 95% confidence interval. So there is a wide range to the bounds / error.