# Maternal Smoking and Birth Weights

*Oblasova Iuliia*

*9/23/2018*

```
smoking <- read.csv("~/Desktop/702 - Data modeling/smoking.csv")
names(smoking)
```

```
##  [1] "id"        "date"      "gestation" "bwt.oz"    "parity"
##  [6] "mrace"     "mage"      "med"       "mht"       "mpregwt"
## [11] "inc"       "smoke"
```

```
dim(smoking)
```

```
## [1] 869  12
```

The new file contains 12 variables, four of which are categorical: mrace (mother's race), med(mother's education), inc (an income) and smoke (0 if mother does not smoke and 1 otherwise).

Also, we make a reasonable assumption that id and date of birth do not effect the weight of the children, so we will not include them into the model. As it was mentioned in the instructions, gestation and weight are "two variables happen simultaneously and hence are a bivariate outcome", so gestation variable is also excluded from the model.

Let's collapse all levels of race from 0 to 5 into one category "white" = 0.

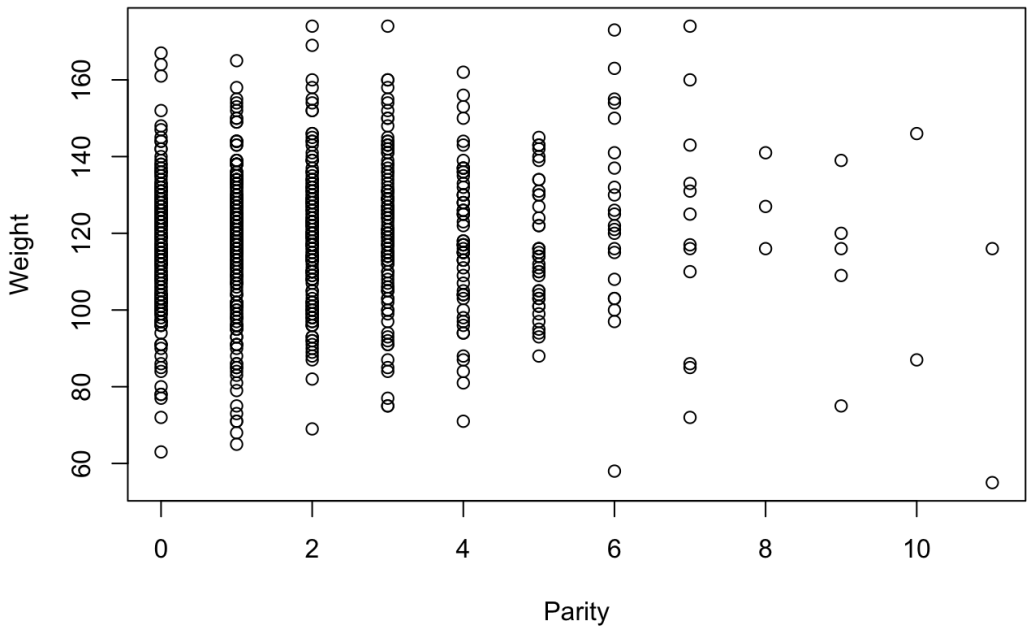
```
smoking$mrace[smoking$mrace==1]<-0
smoking$mrace[smoking$mrace==2]<-0
smoking$mrace[smoking$mrace==3]<-0
smoking$mrace[smoking$mrace==4]<-0
smoking$mrace[smoking$mrace==5]<-0
```

```
summary(smoking)
```

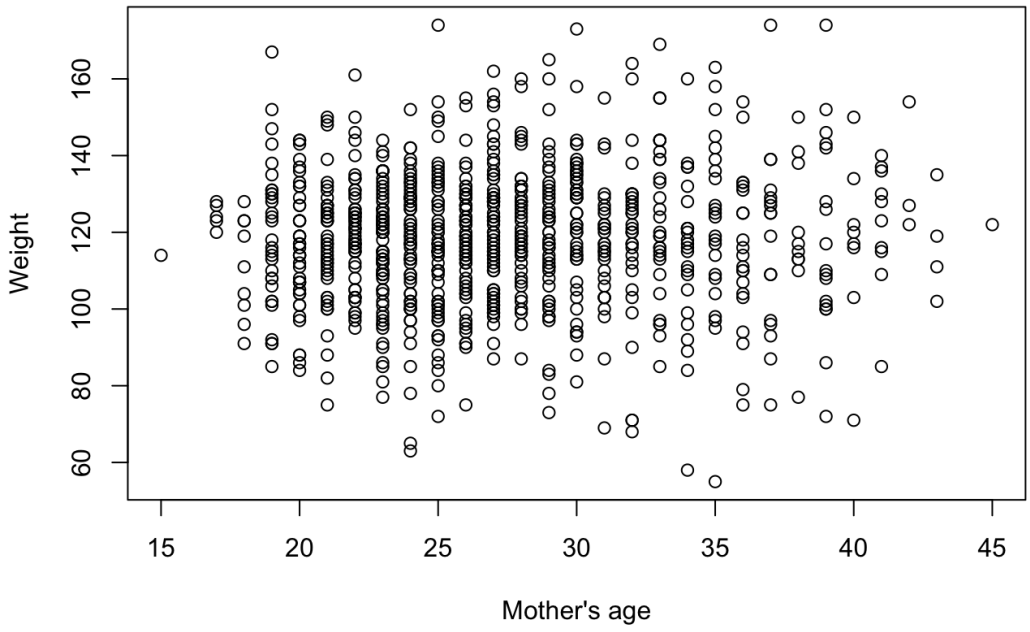
```
##        id            date        gestation         bwt.oz
##  Min.   :  15   Min.   :1350   Min.   :148.0   Min.   : 55.0
##  1st Qu.:5477   1st Qu.:1444   1st Qu.:272.0   1st Qu.:108.0
##  Median :6734   Median :1540   Median :279.0   Median :119.0
##  Mean   :6032   Mean   :1536   Mean   :278.5   Mean   :118.4
##  3rd Qu.:7587   3rd Qu.:1627   3rd Qu.:286.0   3rd Qu.:129.0
##  Max.   :9263   Max.   :1714   Max.   :338.0   Max.   :174.0
##      parity           mrace            mage            med
##  Min.   : 0.000   Min.   :0.000   Min.   :15.00   Min.   :0.000
##  1st Qu.: 1.000   1st Qu.:0.000   1st Qu.:23.00   1st Qu.:2.000
##  Median : 2.000   Median :0.000   Median :26.00   Median :2.000
##  Mean   : 1.953   Mean   :2.002   Mean   :27.29   Mean   :2.932
##  3rd Qu.: 3.000   3rd Qu.:7.000   3rd Qu.:31.00   3rd Qu.:4.000
##  Max.   :11.000   Max.   :9.000   Max.   :45.00   Max.   :7.000
##       mht           mpregwt           inc            smoke
##  Min.   :53.00   Min.   : 87.0   Min.   :0.000   Min.   :0.0000
##  1st Qu.:62.00   1st Qu.:113.0   1st Qu.:2.000   1st Qu.:0.0000
##  Median :64.00   Median :125.0   Median :3.000   Median :0.0000
##  Mean   :64.07   Mean   :128.5   Mean   :3.681   Mean   :0.4638
##  3rd Qu.:66.00   3rd Qu.:140.0   3rd Qu.:5.000   3rd Qu.:1.0000
##  Max.   :72.00   Max.   :220.0   Max.   :9.000   Max.   :1.0000
```

Parity, mpregwt and mht variables may have a few outliers.

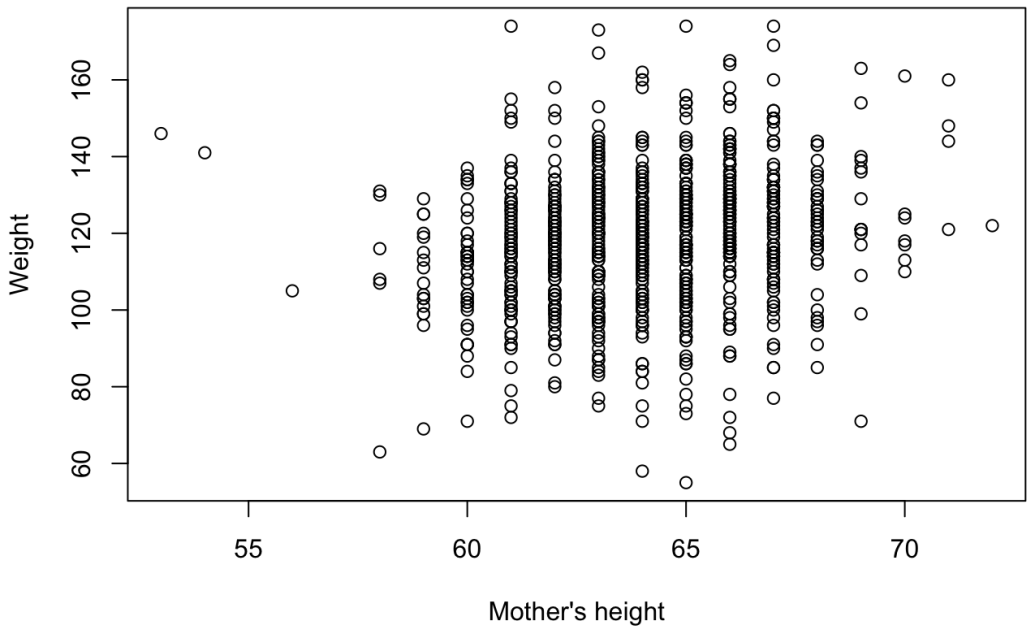
```
plot(smoking$bwt.oz~smoking$parity, xlab = "Parity", ylab = "Weight")
```



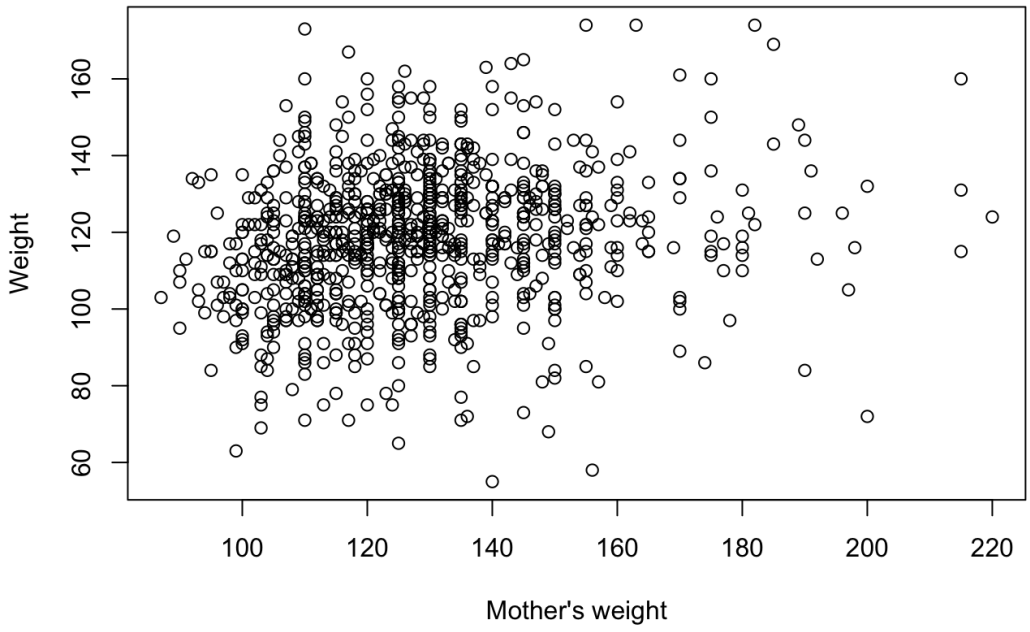

```
plot(smoking$bwt.oz~smoking$mage, xlab = "Mother's age", ylab = "Weight")
```




```
plot(smoking$bwt.oz~smoking$mht, xlab = "Mother's height", ylab = "Weight")
```



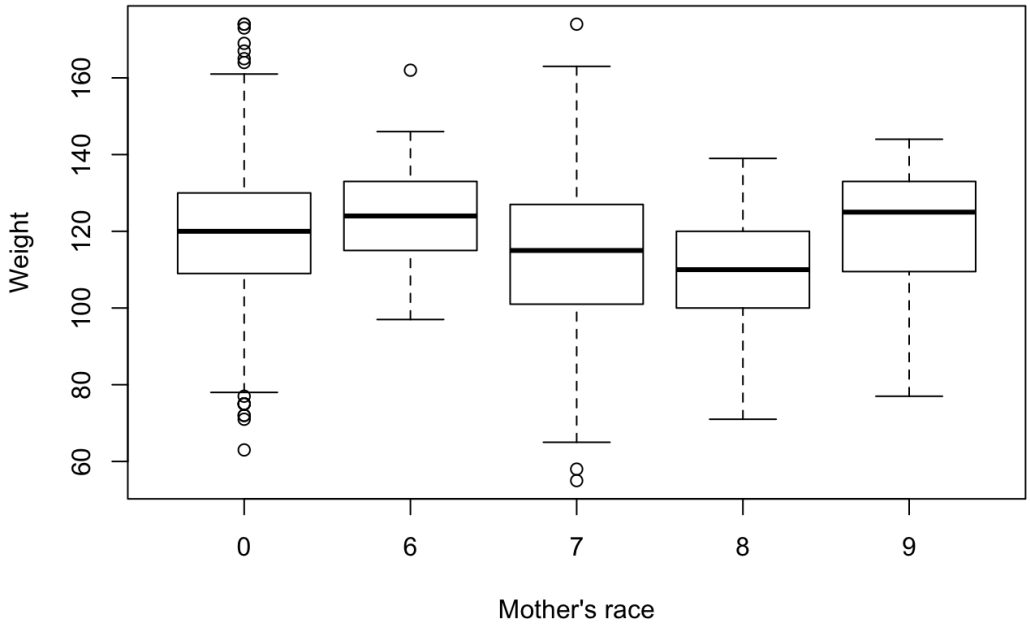

```
plot(smoking$bwt.oz~smoking$mpregwt, xlab = "Mother's weight", ylab = "Weight")
```



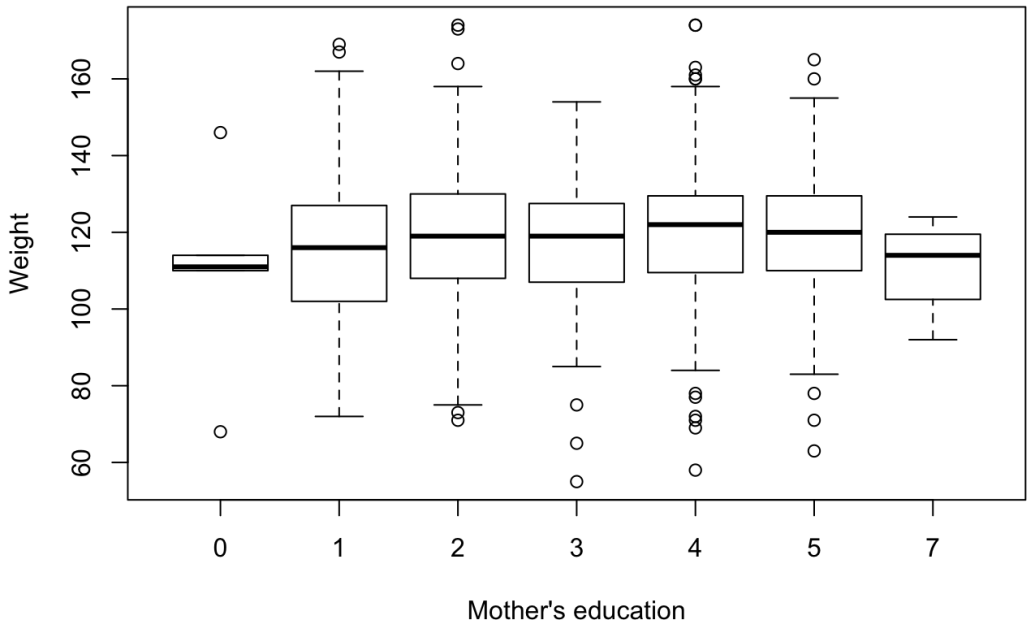

The plots prove the assumption about outliers: mother's height, weight and parity have a few leverage points which may effect the model. Mother's weight variable has the logarifmic shape so we will make a transormation to see if it improves the model.

Let's explore categorical variable now.

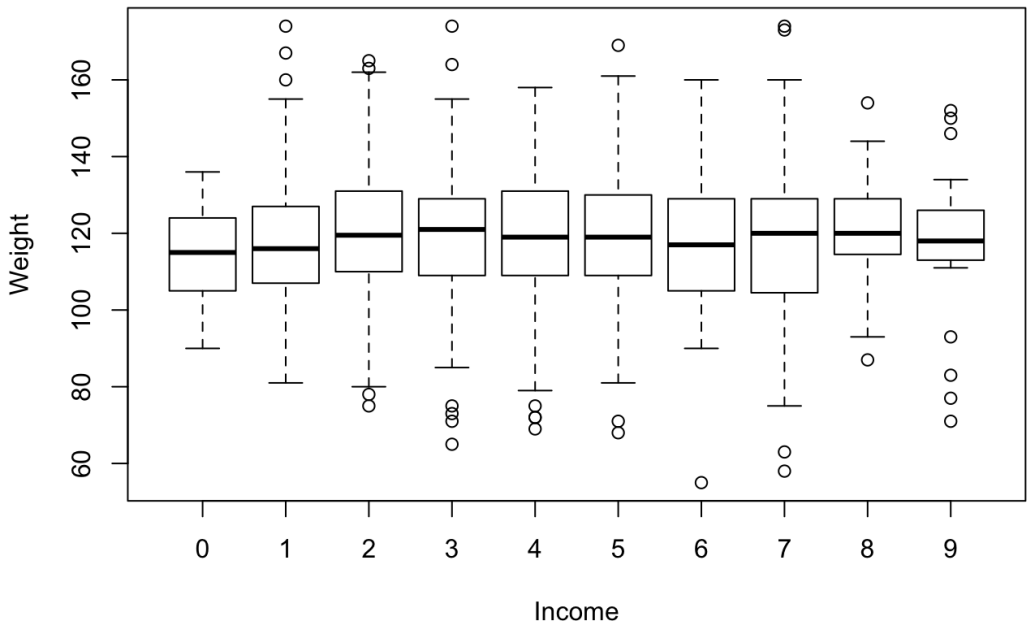
```
boxplot(bwt.oz ~ mrace, data = smoking, xlab = "Mother's race", ylab = "Weight")
```



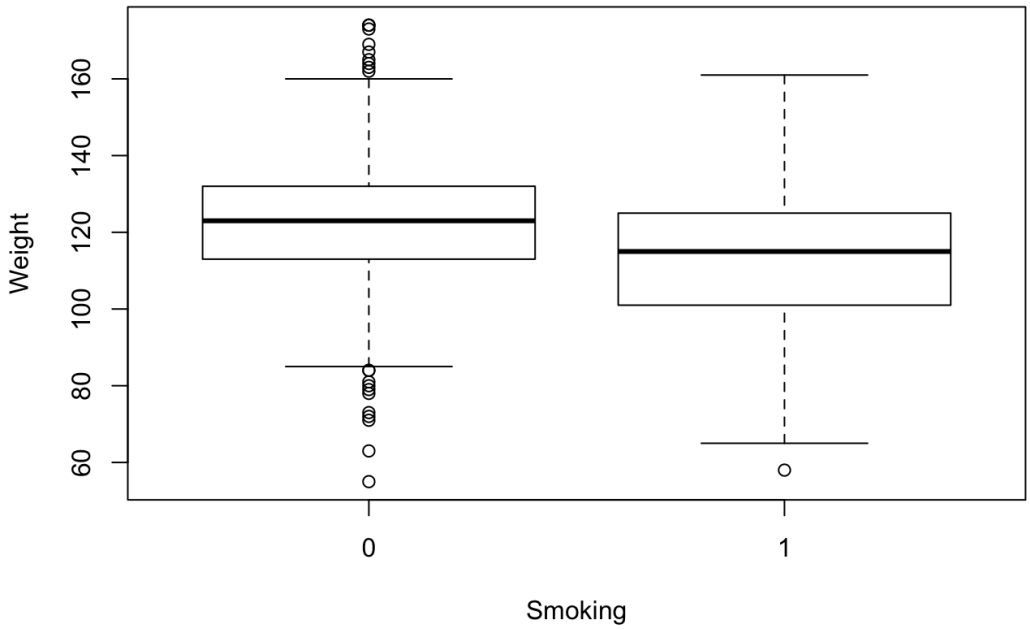

```
boxplot(bwt.oz ~ med, data = smoking, xlab = "Mother's education", ylab = "Weight")
```




```
boxplot(bwt.oz ~ inc, data = smoking, xlab = "Income", ylab = "Weight")
```



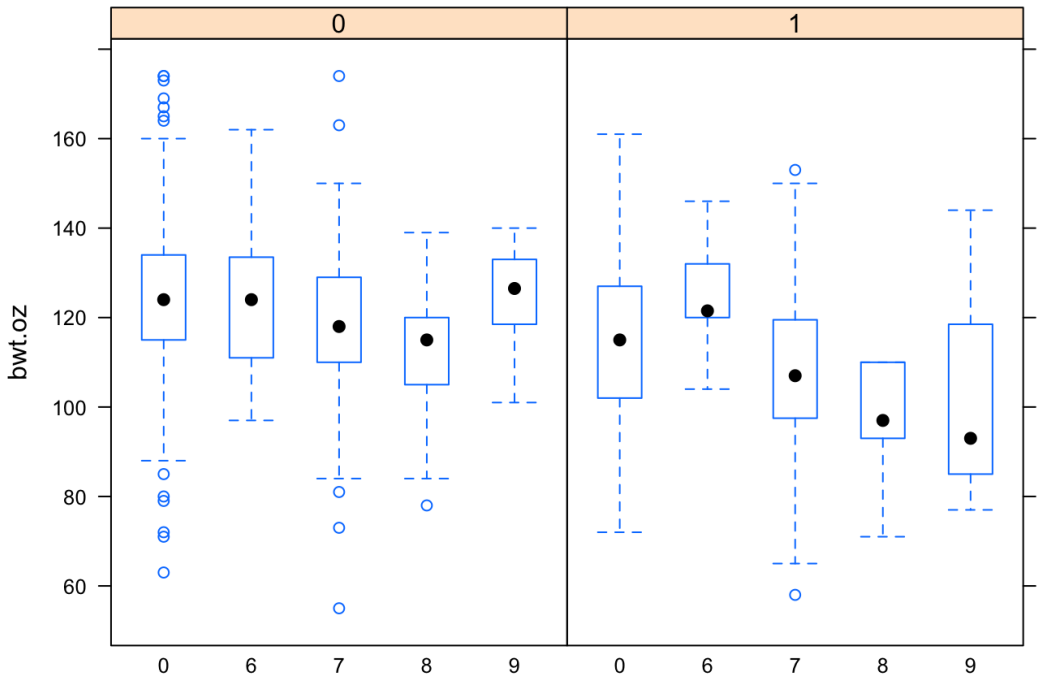

```
boxplot(bwt.oz ~ smoke, data = smoking, xlab = "Smoking", ylab = "Weight")
```



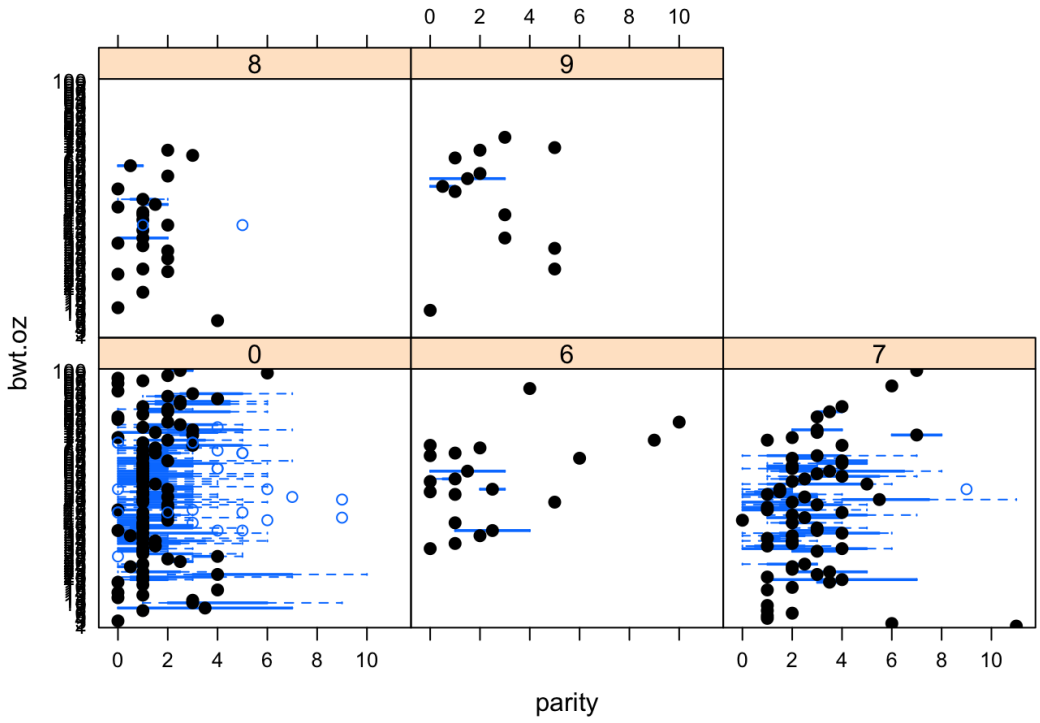

Only smoke variable has an effect on the weight of the children (children of smoking mothers weight less). Also, we do have significantly less information about mothers with 0 level of education (less than 8th grade).

Let's check some reasonable interactions.

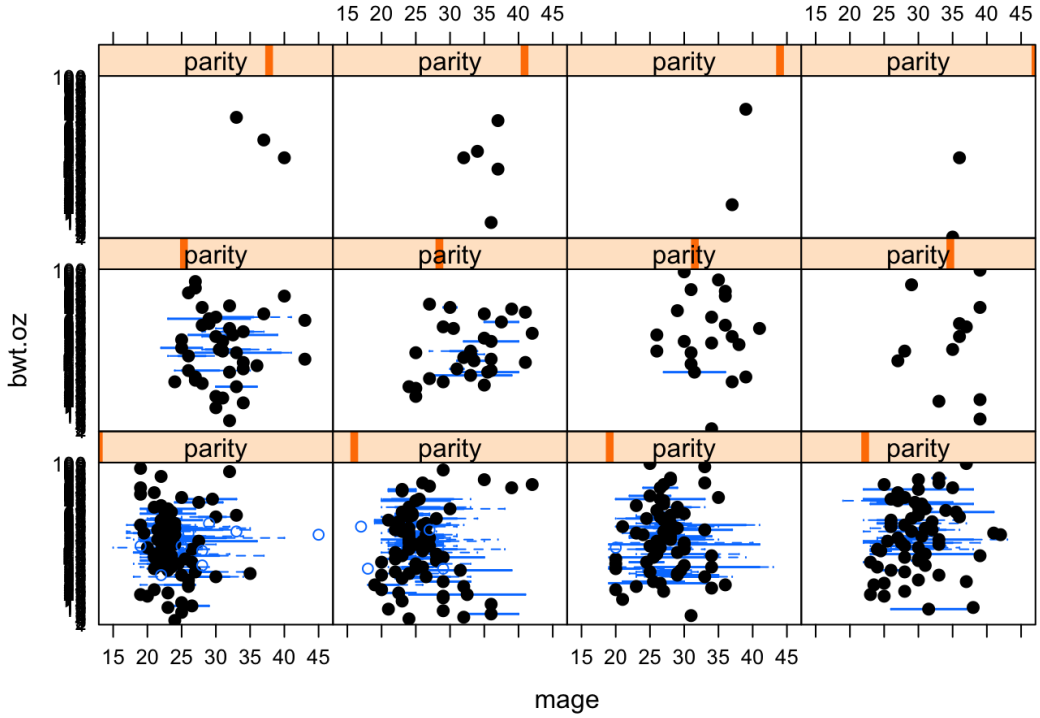
```
library(lattice)
bwplot(bwt.oz ~ as.factor(mrace) | as.factor(smoke), data = smoking)
```



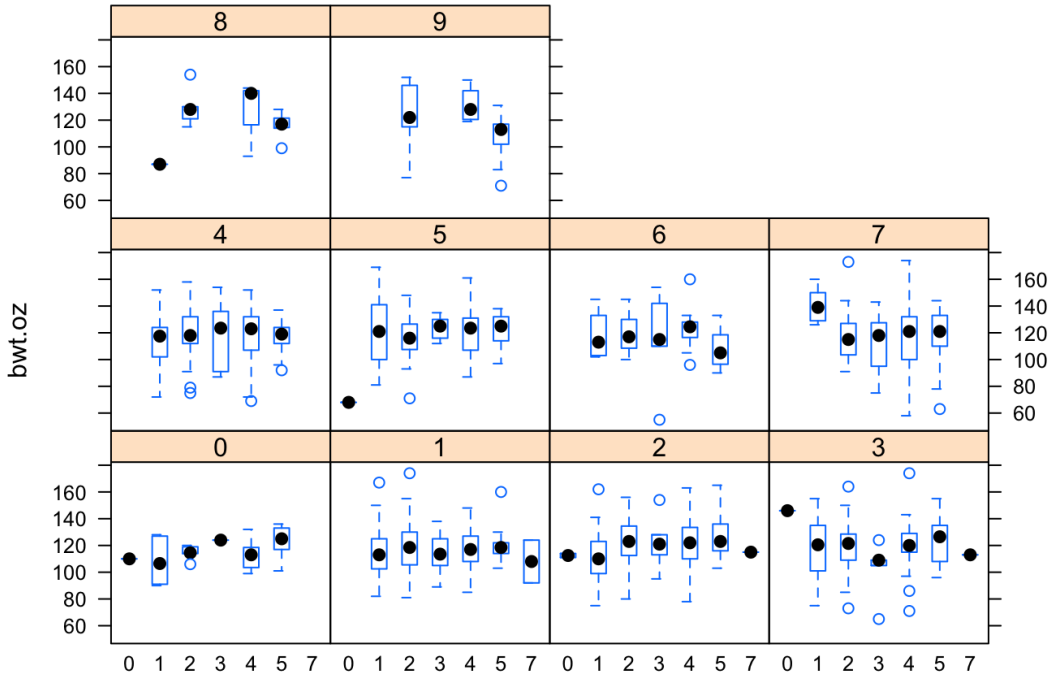

```
bwplot(bwt.oz ~ parity | as.factor(mrace), data = smoking)
```



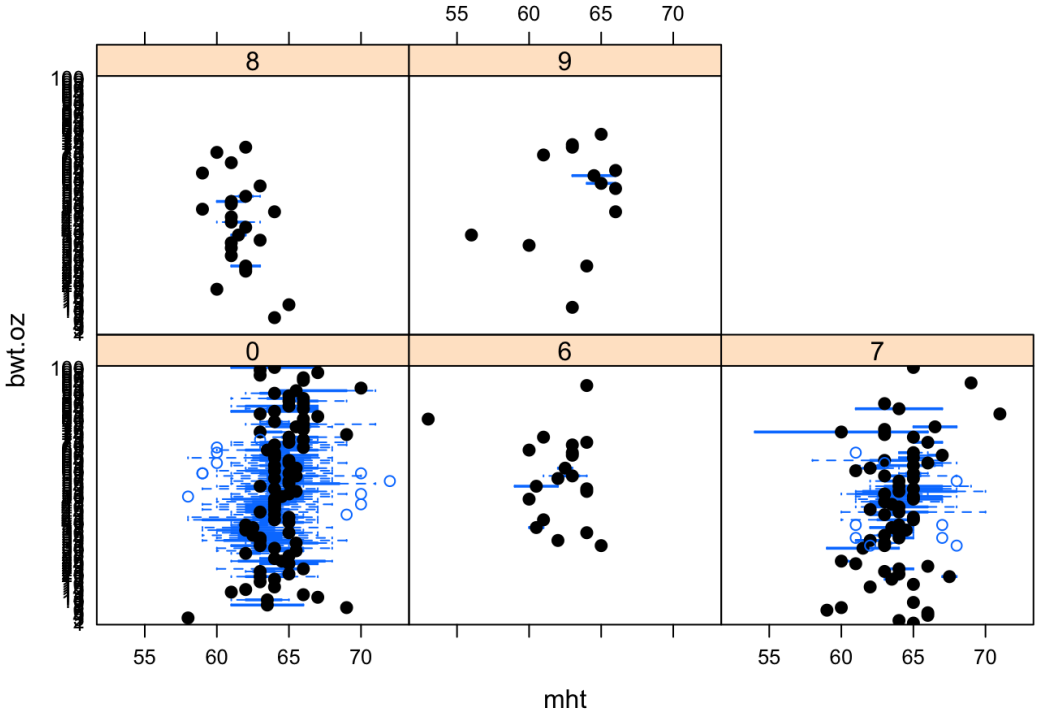

```
bwplot(bwt.oz ~ mage | parity, data = smoking)
```



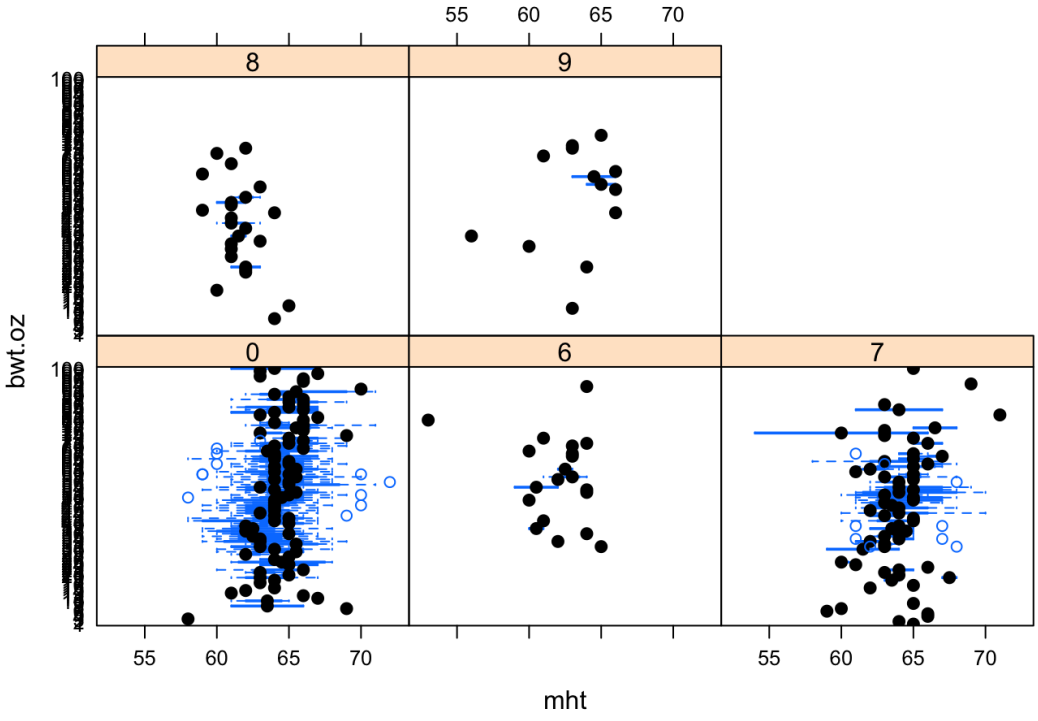

```
bwplot(bwt.oz ~ as.factor(med) | as.factor(inc), data = smoking)
```




```
bwplot(bwt.oz ~ mht | as.factor(mrace), data = smoking)
```




```
bwplot(bwt.oz ~ mht | as.factor(mrace), data = smoking)
```




The first three plots look slightly different, so we include it into the model and see if it makes an improvement.

```
round(cor(smoking[4:12]), 3)
```

```
##         bwt.oz parity  mrace   mage    med    mht mpregwt    inc  smoke
## bwt.oz   1.000  0.041 -0.149  0.044  0.038  0.188   0.182  0.002 -0.249
## parity   0.041  1.000  0.184  0.524 -0.201 -0.043   0.151  0.009  0.011
## mrace   -0.149  0.184  1.000  0.036 -0.051 -0.152   0.061 -0.142 -0.105
## mage     0.044  0.524  0.036  1.000  0.134 -0.005   0.146  0.297 -0.070
## med      0.038 -0.201 -0.051  0.134  1.000  0.115  -0.054  0.217 -0.138
## mht      0.188 -0.043 -0.152 -0.005  0.115  1.000   0.460  0.071  0.041
## mpregwt  0.182  0.151  0.061  0.146 -0.054  0.460   1.000 -0.005 -0.049
## inc      0.002  0.009 -0.142  0.297  0.217  0.071  -0.005  1.000  0.007
## smoke   -0.249  0.011 -0.105 -0.070 -0.138  0.041  -0.049  0.007  1.000
```

As expected, the strongest correlation is between mother's age and gestation and between mother's weight and height.

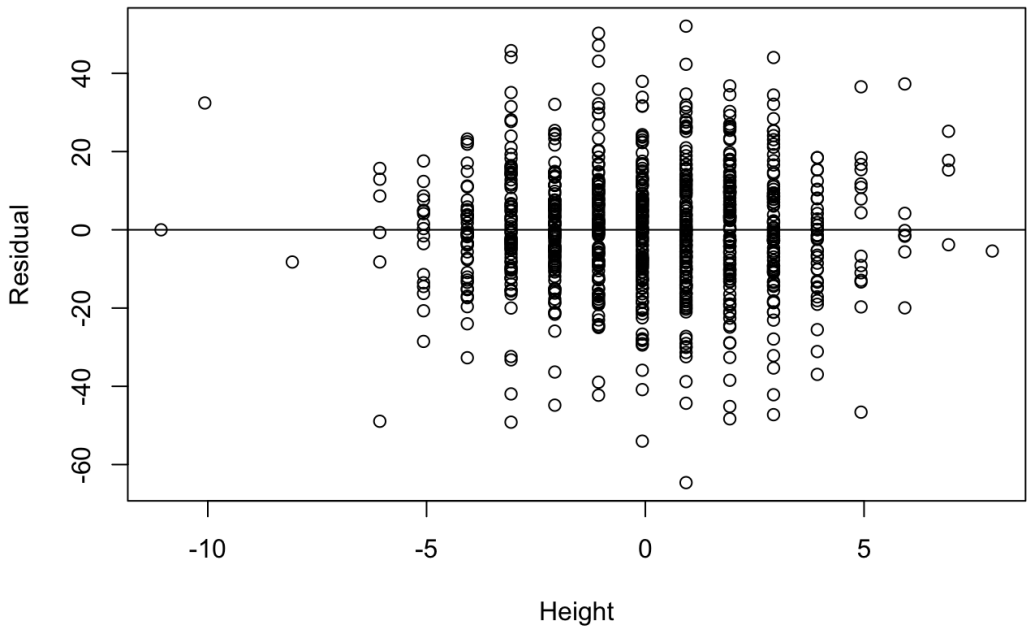
```
#Mean centering.
smoking$mht_cent=smoking$mht-mean(smoking$mht)
smoking$parity_cent=smoking$parity-mean(smoking$parity)
smoking$mpregwt_cent=smoking$mpregwt-mean(smoking$mpregwt)
smoking$mage_cent=smoking$mage-mean(smoking$mage)

wt_reg = lm(bwt.oz~mht+parity+log(mpregwt)+mage+as.factor(inc)+as.factor(mrace)+as.factor(smoke)+as.factor(me
d)+as.factor(med)*as.factor(inc)+as.factor(smoke)*as.factor(mrace)+parity*mage, data=smoking)
summary(wt_reg)
```

```
##
## Call:
## lm(formula = bwt.oz ~ mht + parity + log(mpregwt) + mage + as.factor(inc) +
##     as.factor(mrace) + as.factor(smoke) + as.factor(med) + as.factor(med) *
##     as.factor(inc) + as.factor(smoke) * as.factor(mrace) + parity *
##     mage, data = smoking)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -64.610  -9.625   0.000   9.752  52.042
##
## Coefficients: (16 not defined because of singularities)
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         -40.27927   26.59500  -1.515 0.130283
## mht                                   1.04403    0.28270   3.693 0.000237
## parity                                2.92980    1.78853   1.638 0.101792
## log(mpregwt)                         16.00171    4.59788   3.480 0.000528
## mage                                  0.06644    0.17091   0.389 0.697554
## as.factor(inc)1                      51.21318   32.62980   1.570 0.116921
## as.factor(inc)2                      13.25132   20.34337   0.651 0.514986
## as.factor(inc)3                      50.93134   25.17171   2.023 0.043368
## as.factor(inc)4                      -5.86914    8.52882  -0.688 0.491555
## as.factor(inc)5                     -26.56937   23.53590  -1.129 0.259284
## as.factor(inc)6                     -15.12567    8.67143  -1.744 0.081488
## as.factor(inc)7                      -7.92663    8.02209  -0.988 0.323402
## as.factor(inc)8                     -10.00771    9.82259  -1.019 0.308584
## as.factor(inc)9                     -19.84680    9.11034  -2.178 0.029660
## as.factor(mrace)6                     1.24980    4.06210   0.308 0.758412
## as.factor(mrace)7                    -9.73731    2.15975  -4.509 7.50e-06
## as.factor(mrace)8                    -6.24286    3.68490  -1.694 0.090621
## as.factor(mrace)9                    -0.09601    5.06580  -0.019 0.984883
## as.factor(smoke)1                   -10.18377    1.39136  -7.319 6.07e-13
## as.factor(med)1                      11.04263   18.07015   0.611 0.541308
## as.factor(med)2                      10.84707   18.01021   0.602 0.547163
## as.factor(med)3                      23.79451   23.57086   1.009 0.313045
## as.factor(med)4                      13.46421   17.81685   0.756 0.450051
## as.factor(med)5                      24.45396   18.28581   1.337 0.181498
## as.factor(med)7                     -44.86565   25.30301  -1.773 0.076586
## as.factor(inc)1:as.factor(med)1     -45.32273   33.30572  -1.361 0.173957
## as.factor(inc)2:as.factor(med)1     -11.25339   21.67783  -0.519 0.603821
## as.factor(inc)3:as.factor(med)1     -49.78241   26.28331  -1.894 0.058576
## as.factor(inc)4:as.factor(med)1      10.24286   11.67941   0.877 0.380748
## as.factor(inc)5:as.factor(med)1      35.86383   25.17624   1.425 0.154688
## as.factor(inc)6:as.factor(med)1      22.61584   12.94980   1.746 0.081121
## as.factor(inc)7:as.factor(med)1      39.05983   12.86626   3.036 0.002476
## as.factor(inc)8:as.factor(med)1     -14.79079   20.84512  -0.710 0.478186
## as.factor(inc)9:as.factor(med)1            NA         NA      NA       NA
## as.factor(inc)1:as.factor(med)2     -43.16022   33.35239  -1.294 0.196015
## as.factor(inc)2:as.factor(med)2      -2.65235   21.55556  -0.123 0.902101
## as.factor(inc)3:as.factor(med)2     -44.72918   26.19503  -1.708 0.088108
## as.factor(inc)4:as.factor(med)2      14.87348   11.21418   1.326 0.185115
## as.factor(inc)5:as.factor(med)2      31.24113   24.65436   1.267 0.205465
## as.factor(inc)6:as.factor(med)2      21.30803   11.82035   1.803 0.071818
## as.factor(inc)7:as.factor(med)2      12.34220   10.81064   1.142 0.253932
## as.factor(inc)8:as.factor(med)2      27.33779   14.03316   1.948 0.051754
## as.factor(inc)9:as.factor(med)2      32.04526   13.18022   2.431 0.015262
## as.factor(inc)1:as.factor(med)3     -61.40351   37.23923  -1.649 0.099562
## as.factor(inc)2:as.factor(med)3     -13.42780   27.18879  -0.494 0.621532
## as.factor(inc)3:as.factor(med)3     -68.89922   30.98472  -2.224 0.026450
## as.factor(inc)4:as.factor(med)3      -2.41095   19.91073  -0.121 0.903652
## as.factor(inc)5:as.factor(med)3      23.50718   29.70323   0.791 0.428944
## as.factor(inc)6:as.factor(med)3       6.82227   20.27277   0.337 0.736564
## as.factor(inc)7:as.factor(med)3       0.14150   19.25073   0.007 0.994137
## as.factor(inc)8:as.factor(med)3            NA         NA      NA       NA
## as.factor(inc)9:as.factor(med)3            NA         NA      NA       NA
## as.factor(inc)1:as.factor(med)4     -47.22914   33.41819  -1.413 0.157963
```

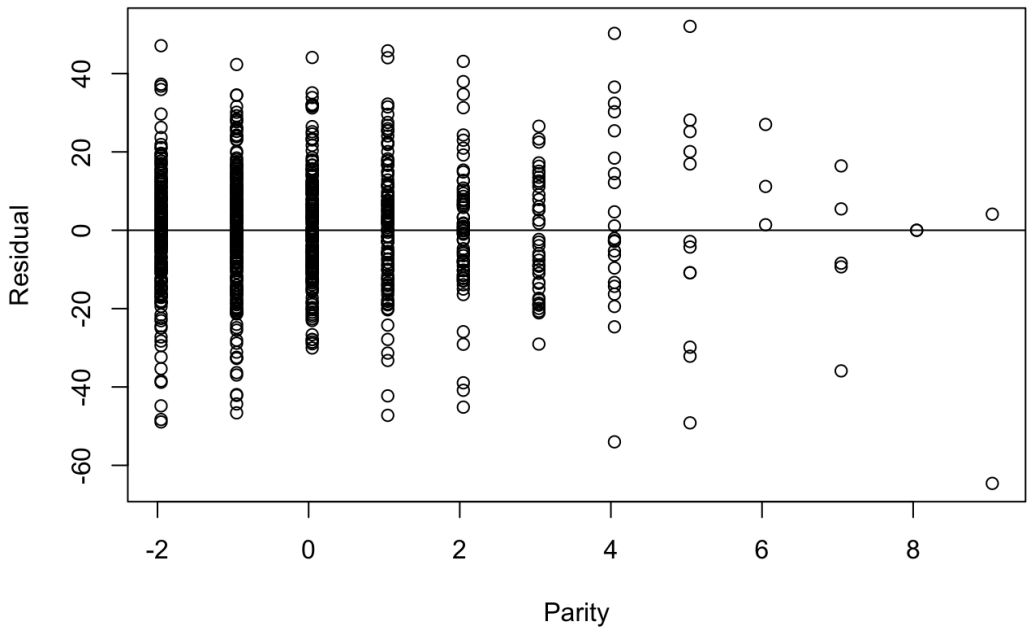
```
## as.factor(inc)2:as.factor(med)4      -6.23318   21.54823  -0.289 0.772452
## as.factor(inc)3:as.factor(med)4     -45.13855   26.09789  -1.730 0.084089
## as.factor(inc)4:as.factor(med)4       7.84378   11.03381   0.711 0.477362
## as.factor(inc)5:as.factor(med)4      30.56120   24.66010   1.239 0.215598
## as.factor(inc)6:as.factor(med)4      22.16083   11.50387   1.926 0.054410
## as.factor(inc)7:as.factor(med)4      10.69185   10.82814   0.987 0.323738
## as.factor(inc)8:as.factor(med)4      18.89852   15.20781   1.243 0.214348
## as.factor(inc)9:as.factor(med)4      32.17001   13.80338   2.331 0.020022
## as.factor(inc)1:as.factor(med)5     -56.43695   33.81228  -1.669 0.095483
## as.factor(inc)2:as.factor(med)5     -12.98698   22.09808  -0.588 0.556901
## as.factor(inc)3:as.factor(med)5     -58.24962   26.64757  -2.186 0.029110
## as.factor(inc)4:as.factor(med)5            NA         NA      NA       NA
## as.factor(inc)5:as.factor(med)5      22.53299   25.04874   0.900 0.368622
## as.factor(inc)6:as.factor(med)5            NA         NA      NA       NA
## as.factor(inc)7:as.factor(med)5            NA         NA      NA       NA
## as.factor(inc)8:as.factor(med)5            NA         NA      NA       NA
## as.factor(inc)9:as.factor(med)5            NA         NA      NA       NA
## as.factor(inc)1:as.factor(med)7            NA         NA      NA       NA
## as.factor(inc)2:as.factor(med)7      38.21720   32.49971   1.176 0.239975
## as.factor(inc)3:as.factor(med)7            NA         NA      NA       NA
## as.factor(inc)4:as.factor(med)7            NA         NA      NA       NA
## as.factor(inc)5:as.factor(med)7            NA         NA      NA       NA
## as.factor(inc)6:as.factor(med)7            NA         NA      NA       NA
## as.factor(inc)7:as.factor(med)7            NA         NA      NA       NA
## as.factor(inc)8:as.factor(med)7            NA         NA      NA       NA
## as.factor(inc)9:as.factor(med)7            NA         NA      NA       NA
## as.factor(mrace)6:as.factor(smoke)1   9.55264    8.59985   1.111 0.266991
## as.factor(mrace)7:as.factor(smoke)1   3.00952    3.02000   0.997 0.319295
## as.factor(mrace)8:as.factor(smoke)1  -5.50192    6.73728  -0.817 0.414378
## as.factor(mrace)9:as.factor(smoke)1 -11.79938   11.01512  -1.071 0.284403
## parity:mage                          -0.07105    0.05673  -1.252 0.210801
##
## (Intercept)
## mht                                 ***
## parity
## log(mpregwt)                        ***
## mage
## as.factor(inc)1
## as.factor(inc)2
## as.factor(inc)3                     *
## as.factor(inc)4
## as.factor(inc)5
## as.factor(inc)6                     .
## as.factor(inc)7
## as.factor(inc)8
## as.factor(inc)9                     *
## as.factor(mrace)6
## as.factor(mrace)7                   ***
## as.factor(mrace)8                   .
## as.factor(mrace)9
## as.factor(smoke)1                   ***
## as.factor(med)1
## as.factor(med)2
## as.factor(med)3
## as.factor(med)4
## as.factor(med)5
## as.factor(med)7                     .
## as.factor(inc)1:as.factor(med)1
## as.factor(inc)2:as.factor(med)1
## as.factor(inc)3:as.factor(med)1     .
## as.factor(inc)4:as.factor(med)1
## as.factor(inc)5:as.factor(med)1
## as.factor(inc)6:as.factor(med)1     .
## as.factor(inc)7:as.factor(med)1     **
## as.factor(inc)8:as.factor(med)1
## as.factor(inc)9:as.factor(med)1
## as.factor(inc)1:as.factor(med)2
```

```
## as.factor(inc)2:as.factor(med)2
## as.factor(inc)3:as.factor(med)2     .
## as.factor(inc)4:as.factor(med)2
## as.factor(inc)5:as.factor(med)2
## as.factor(inc)6:as.factor(med)2     .
## as.factor(inc)7:as.factor(med)2
## as.factor(inc)8:as.factor(med)2     .
## as.factor(inc)9:as.factor(med)2     *
## as.factor(inc)1:as.factor(med)3     .
## as.factor(inc)2:as.factor(med)3
## as.factor(inc)3:as.factor(med)3     *
## as.factor(inc)4:as.factor(med)3
## as.factor(inc)5:as.factor(med)3
## as.factor(inc)6:as.factor(med)3
## as.factor(inc)7:as.factor(med)3
## as.factor(inc)8:as.factor(med)3
## as.factor(inc)9:as.factor(med)3
## as.factor(inc)1:as.factor(med)4
## as.factor(inc)2:as.factor(med)4
## as.factor(inc)3:as.factor(med)4     .
## as.factor(inc)4:as.factor(med)4
## as.factor(inc)5:as.factor(med)4
## as.factor(inc)6:as.factor(med)4     .
## as.factor(inc)7:as.factor(med)4
## as.factor(inc)8:as.factor(med)4
## as.factor(inc)9:as.factor(med)4     *
## as.factor(inc)1:as.factor(med)5     .
## as.factor(inc)2:as.factor(med)5
## as.factor(inc)3:as.factor(med)5     *
## as.factor(inc)4:as.factor(med)5
## as.factor(inc)5:as.factor(med)5
## as.factor(inc)6:as.factor(med)5
## as.factor(inc)7:as.factor(med)5
## as.factor(inc)8:as.factor(med)5
## as.factor(inc)9:as.factor(med)5
## as.factor(inc)1:as.factor(med)7
## as.factor(inc)2:as.factor(med)7
## as.factor(inc)3:as.factor(med)7
## as.factor(inc)4:as.factor(med)7
## as.factor(inc)5:as.factor(med)7
## as.factor(inc)6:as.factor(med)7
## as.factor(inc)7:as.factor(med)7
## as.factor(inc)8:as.factor(med)7
## as.factor(inc)9:as.factor(med)7
## as.factor(mrace)6:as.factor(smoke)1
## as.factor(mrace)7:as.factor(smoke)1
## as.factor(mrace)8:as.factor(smoke)1
## as.factor(mrace)9:as.factor(smoke)1
## parity:mage
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.54 on 801 degrees of freedom
## Multiple R-squared:  0.2256, Adjusted R-squared:  0.1609
## F-statistic: 3.484 on 67 and 801 DF,  p-value: < 2.2e-16
```

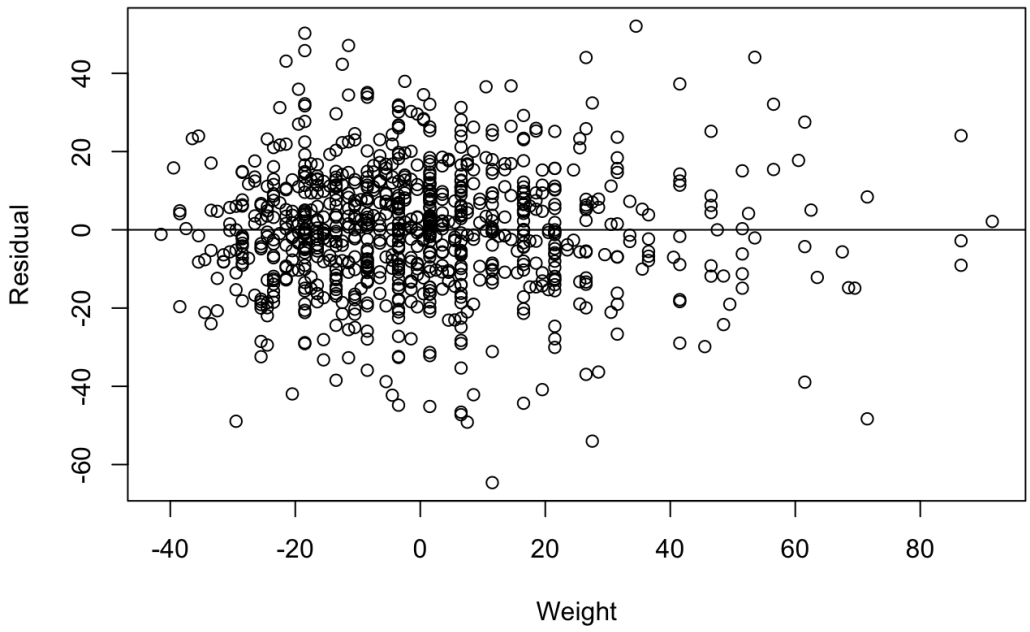
```
plot(y = wt_reg$resid, x = smoking$mht_cent, xlab = "Height", ylab = "Residual")
abline(0,0)
```



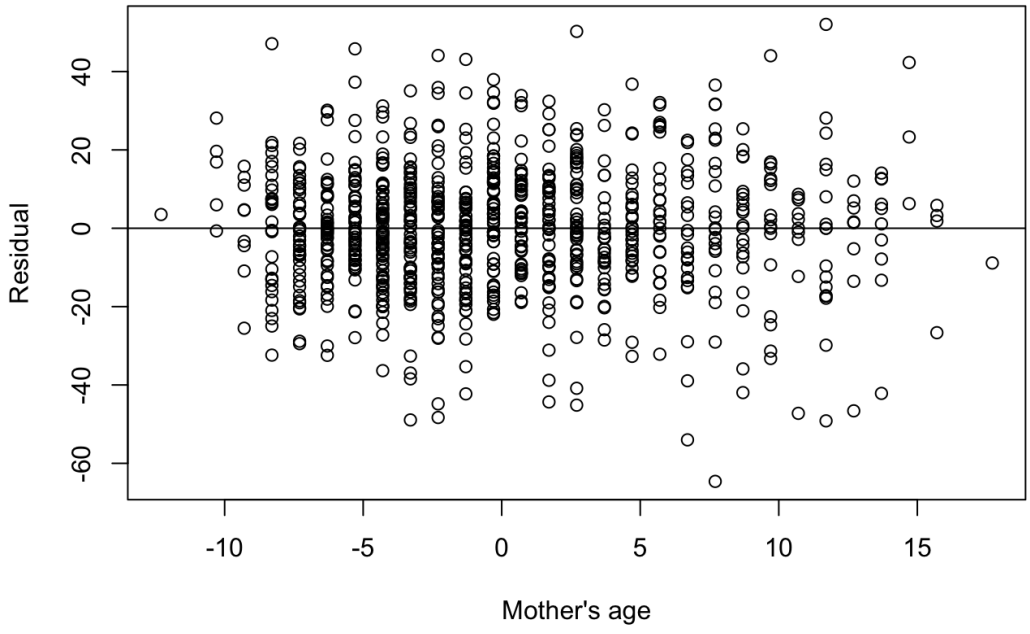

```
plot(y = wt_reg$resid, x = smoking$parity_cent, xlab = "Parity", ylab = "Residual")
abline(0,0)
```




```
plot(y = wt_reg$resid, x = smoking$mpregwt_cent, xlab = "Weight", ylab = "Residual")
abline(0,0)
```




```
plot(y = wt_reg$resid, x = smoking$mage_cent, xlab = "Mother's age", ylab = "Residual")
abline(0,0)
```




Residual plots do not show any patterns, so we assume this is a good model. However, height variable has leverage points, so we remove it to see if it improves the model and test the model without log transformation of mpregwt varibale.

```
corrected_height <- smoking[smoking$mht > 57,]
wt_reg_2= lm(bwt.oz~mht+parity+mpregwt+mage+as.factor(inc)+as.factor(mrace)+as.factor(smoke)+as.factor(med)+a
s.factor(med)*as.factor(inc)+as.factor(smoke)*as.factor(mrace)+parity*mage, data=corrected_height)
summary(wt_reg_2)
```

```
##
## Call:
## lm(formula = bwt.oz ~ mht + parity + mpregwt + mage + as.factor(inc) +
##     as.factor(mrace) + as.factor(smoke) + as.factor(med) + as.factor(med) *
##     as.factor(inc) + as.factor(smoke) * as.factor(mrace) + parity *
##     mage, data = corrected_height)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -64.185  -9.742   0.000   9.755  52.123
##
## Coefficients: (17 not defined because of singularities)
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         14.11324   24.70762   0.571  0.56802
## mht                                  1.18986    0.28291   4.206 2.90e-05
## parity                               2.76769    1.79513   1.542  0.12352
## mpregwt                              0.10207    0.03376   3.023  0.00258
## mage                                 0.07747    0.17073   0.454  0.65011
## as.factor(inc)1                     13.19736   28.79728   0.458  0.64687
## as.factor(inc)2                     13.91811   20.32695   0.685  0.49372
## as.factor(inc)3                     12.64976   31.05461   0.407  0.68387
## as.factor(inc)4                     -5.80172    8.52223  -0.681  0.49621
## as.factor(inc)5                    -25.67262   23.51951  -1.092  0.27536
## as.factor(inc)6                    -15.06302    8.66508  -1.738  0.08253
## as.factor(inc)7                     -7.93539    8.01584  -0.990  0.32249
## as.factor(inc)8                     -9.75300    9.81412  -0.994  0.32063
## as.factor(inc)9                    -19.89188    9.10336  -2.185  0.02917
## as.factor(mrace)6                    1.19319    4.05893   0.294  0.76886
## as.factor(mrace)7                   -9.59033    2.16089  -4.438 1.03e-05
## as.factor(mrace)8                   -6.43204    3.67932  -1.748  0.08082
## as.factor(mrace)9                    0.80165    5.28553   0.152  0.87949
## as.factor(smoke)1                  -10.30407    1.38935  -7.416 3.07e-13
## as.factor(med)1                     11.63362   18.05993   0.644  0.51965
## as.factor(med)2                     11.24740   18.00513   0.625  0.53236
## as.factor(med)3                     24.41591   23.56504   1.036  0.30047
## as.factor(med)4                     13.94133   17.80373   0.783  0.43383
## as.factor(med)5                     24.88127   18.27289   1.362  0.17369
## as.factor(med)7                     -6.11938   20.42148  -0.300  0.76452
## as.factor(inc)1:as.factor(med)1     -7.11781   29.68431  -0.240  0.81056
## as.factor(inc)2:as.factor(med)1    -11.84073   21.66101  -0.547  0.58478
## as.factor(inc)3:as.factor(med)1    -11.58071   32.12122  -0.361  0.71854
## as.factor(inc)4:as.factor(med)1     10.52984   11.66910   0.902  0.36713
## as.factor(inc)5:as.factor(med)1     34.92725   25.15971   1.388  0.16546
## as.factor(inc)6:as.factor(med)1     22.64726   12.93972   1.750  0.08047
## as.factor(inc)7:as.factor(med)1     38.70207   12.85712   3.010  0.00269
## as.factor(inc)8:as.factor(med)1    -14.80576   20.82981  -0.711  0.47742
## as.factor(inc)9:as.factor(med)1           NA         NA      NA       NA
## as.factor(inc)1:as.factor(med)2     -4.99079   29.67251  -0.168  0.86647
## as.factor(inc)2:as.factor(med)2     -3.64737   21.54572  -0.169  0.86562
## as.factor(inc)3:as.factor(med)2     -6.16152   31.89405  -0.193  0.84686
## as.factor(inc)4:as.factor(med)2     15.11338   11.21143   1.348  0.17803
## as.factor(inc)5:as.factor(med)2     30.44793   24.64473   1.235  0.21702
## as.factor(inc)6:as.factor(med)2     21.16112   11.81604   1.791  0.07369
## as.factor(inc)7:as.factor(med)2     12.45248   10.80657   1.152  0.24954
## as.factor(inc)8:as.factor(med)2     27.47830   14.02706   1.959  0.05047
## as.factor(inc)9:as.factor(med)2     32.54250   13.17297   2.470  0.01370
## as.factor(inc)1:as.factor(med)3    -23.47910   33.94451  -0.692  0.48933
## as.factor(inc)2:as.factor(med)3    -14.19022   27.17767  -0.522  0.60173
## as.factor(inc)3:as.factor(med)3    -30.74312   36.01064  -0.854  0.39351
## as.factor(inc)4:as.factor(med)3     -2.38823   19.89714  -0.120  0.90449
## as.factor(inc)5:as.factor(med)3     22.29085   29.69114   0.751  0.45302
## as.factor(inc)6:as.factor(med)3      6.57483   20.26495   0.324  0.74569
## as.factor(inc)7:as.factor(med)3      0.18411   19.24342   0.010  0.99237
## as.factor(inc)8:as.factor(med)3           NA         NA      NA       NA
## as.factor(inc)9:as.factor(med)3           NA         NA      NA       NA
## as.factor(inc)1:as.factor(med)4     -9.43465   29.63905  -0.318  0.75033
```

```
## as.factor(inc)2:as.factor(med)4      -6.91731   21.52911  -0.321  0.74807
## as.factor(inc)3:as.factor(med)4      -6.84211   31.82690  -0.215  0.82984
## as.factor(inc)4:as.factor(med)4       7.72524   11.02525   0.701  0.48370
## as.factor(inc)5:as.factor(med)4      29.45669   24.64115   1.195  0.23227
## as.factor(inc)6:as.factor(med)4      22.00640   11.49492   1.914  0.05592
## as.factor(inc)7:as.factor(med)4      10.72338   10.82106   0.991  0.32200
## as.factor(inc)8:as.factor(med)4      18.19611   15.19989   1.197  0.23161
## as.factor(inc)9:as.factor(med)4      32.01080   13.79318   2.321  0.02055
## as.factor(inc)1:as.factor(med)5     -18.39733   30.11377  -0.611  0.54142
## as.factor(inc)2:as.factor(med)5     -13.64063   22.08101  -0.618  0.53691
## as.factor(inc)3:as.factor(med)5     -20.04261   32.16558  -0.623  0.53339
## as.factor(inc)4:as.factor(med)5            NA         NA      NA       NA
## as.factor(inc)5:as.factor(med)5      21.72285   25.03233   0.868  0.38577
## as.factor(inc)6:as.factor(med)5            NA         NA      NA       NA
## as.factor(inc)7:as.factor(med)5            NA         NA      NA       NA
## as.factor(inc)8:as.factor(med)5            NA         NA      NA       NA
## as.factor(inc)9:as.factor(med)5            NA         NA      NA       NA
## as.factor(inc)1:as.factor(med)7            NA         NA      NA       NA
## as.factor(inc)2:as.factor(med)7            NA         NA      NA       NA
## as.factor(inc)3:as.factor(med)7            NA         NA      NA       NA
## as.factor(inc)4:as.factor(med)7            NA         NA      NA       NA
## as.factor(inc)5:as.factor(med)7            NA         NA      NA       NA
## as.factor(inc)6:as.factor(med)7            NA         NA      NA       NA
## as.factor(inc)7:as.factor(med)7            NA         NA      NA       NA
## as.factor(inc)8:as.factor(med)7            NA         NA      NA       NA
## as.factor(inc)9:as.factor(med)7            NA         NA      NA       NA
## as.factor(mrace)6:as.factor(smoke)1   9.69524    8.59319   1.128  0.25955
## as.factor(mrace)7:as.factor(smoke)1   2.63303    3.02707   0.870  0.38465
## as.factor(mrace)8:as.factor(smoke)1  -5.51731    6.73202  -0.820  0.41271
## as.factor(mrace)9:as.factor(smoke)1 -12.51653   11.12285  -1.125  0.26080
## parity:mage                          -0.06767    0.05686  -1.190  0.23437
##
## (Intercept)
## mht                                 ***
## parity
## mpregwt                             **
## mage
## as.factor(inc)1
## as.factor(inc)2
## as.factor(inc)3
## as.factor(inc)4
## as.factor(inc)5
## as.factor(inc)6                     .
## as.factor(inc)7
## as.factor(inc)8
## as.factor(inc)9                     *
## as.factor(mrace)6
## as.factor(mrace)7                   ***
## as.factor(mrace)8                   .
## as.factor(mrace)9
## as.factor(smoke)1                   ***
## as.factor(med)1
## as.factor(med)2
## as.factor(med)3
## as.factor(med)4
## as.factor(med)5
## as.factor(med)7
## as.factor(inc)1:as.factor(med)1
## as.factor(inc)2:as.factor(med)1
## as.factor(inc)3:as.factor(med)1
## as.factor(inc)4:as.factor(med)1
## as.factor(inc)5:as.factor(med)1
## as.factor(inc)6:as.factor(med)1     .
## as.factor(inc)7:as.factor(med)1     **
## as.factor(inc)8:as.factor(med)1
## as.factor(inc)9:as.factor(med)1
## as.factor(inc)1:as.factor(med)2
```

```
## as.factor(inc)2:as.factor(med)2
## as.factor(inc)3:as.factor(med)2
## as.factor(inc)4:as.factor(med)2
## as.factor(inc)5:as.factor(med)2
## as.factor(inc)6:as.factor(med)2     .
## as.factor(inc)7:as.factor(med)2
## as.factor(inc)8:as.factor(med)2     .
## as.factor(inc)9:as.factor(med)2     *
## as.factor(inc)1:as.factor(med)3
## as.factor(inc)2:as.factor(med)3
## as.factor(inc)3:as.factor(med)3
## as.factor(inc)4:as.factor(med)3
## as.factor(inc)5:as.factor(med)3
## as.factor(inc)6:as.factor(med)3
## as.factor(inc)7:as.factor(med)3
## as.factor(inc)8:as.factor(med)3
## as.factor(inc)9:as.factor(med)3
## as.factor(inc)1:as.factor(med)4
## as.factor(inc)2:as.factor(med)4
## as.factor(inc)3:as.factor(med)4
## as.factor(inc)4:as.factor(med)4
## as.factor(inc)5:as.factor(med)4
## as.factor(inc)6:as.factor(med)4     .
## as.factor(inc)7:as.factor(med)4
## as.factor(inc)8:as.factor(med)4
## as.factor(inc)9:as.factor(med)4     *
## as.factor(inc)1:as.factor(med)5
## as.factor(inc)2:as.factor(med)5
## as.factor(inc)3:as.factor(med)5
## as.factor(inc)4:as.factor(med)5
## as.factor(inc)5:as.factor(med)5
## as.factor(inc)6:as.factor(med)5
## as.factor(inc)7:as.factor(med)5
## as.factor(inc)8:as.factor(med)5
## as.factor(inc)9:as.factor(med)5
## as.factor(inc)1:as.factor(med)7
## as.factor(inc)2:as.factor(med)7
## as.factor(inc)3:as.factor(med)7
## as.factor(inc)4:as.factor(med)7
## as.factor(inc)5:as.factor(med)7
## as.factor(inc)6:as.factor(med)7
## as.factor(inc)7:as.factor(med)7
## as.factor(inc)8:as.factor(med)7
## as.factor(inc)9:as.factor(med)7
## as.factor(mrace)6:as.factor(smoke)1
## as.factor(mrace)7:as.factor(smoke)1
## as.factor(mrace)8:as.factor(smoke)1
## as.factor(mrace)9:as.factor(smoke)1
## parity:mage
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.52 on 799 degrees of freedom
## Multiple R-squared:  0.2248, Adjusted R-squared:  0.1608
## F-statistic: 3.511 on 66 and 799 DF,  p-value: < 2.2e-16
```

R-squared differs only by 0.3%, so is not reasonable to apply log transformations or remove leverage points.

Let's see if the interactions are important predictors.

```
wt_reg_3 = lm(bwt.oz~mht+parity+mpregwt+mage+as.factor(inc)+as.factor(mrace)+as.factor(smoke)+as.factor(med)+a
s.factor(smoke)*as.factor(mrace)+parity*mage, data=smoking)
anova(wt_reg_3, wt_reg)
```

```
## Analysis of Variance Table
##
## Model 1: bwt.oz ~ mht + parity + mpregwt + mage + as.factor(inc) + as.factor(mrace) +
##     as.factor(smoke) + as.factor(med) + as.factor(smoke) * as.factor(mrace) +
##     parity * mage
## Model 2: bwt.oz ~ mht + parity + log(mpregwt) + mage + as.factor(inc) +
##     as.factor(mrace) + as.factor(smoke) + as.factor(med) + as.factor(med) *
##     as.factor(inc) + as.factor(smoke) * as.factor(mrace) + parity *
##     mage
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    839 234425
## 2    801 219002 38     15423 1.4844 0.03171 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
wt_reg_4=lm(bwt.oz~mht+parity+mpregwt+mage+as.factor(inc)+as.factor(mrace)+as.factor(smoke)+as.factor(med)+as.
factor(med)*as.factor(inc)+parity*mage, data=smoking)
anova(wt_reg_4, wt_reg)
```

```
## Analysis of Variance Table
##
## Model 1: bwt.oz ~ mht + parity + mpregwt + mage + as.factor(inc) + as.factor(mrace) +
##     as.factor(smoke) + as.factor(med) + as.factor(med) * as.factor(inc) +
##     parity * mage
## Model 2: bwt.oz ~ mht + parity + log(mpregwt) + mage + as.factor(inc) +
##     as.factor(mrace) + as.factor(smoke) + as.factor(med) + as.factor(med) *
##     as.factor(inc) + as.factor(smoke) * as.factor(mrace) + parity *
##     mage
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    805 220562
## 2    801 219002  4    1560.5 1.4269 0.2232
```

```
wt_reg_5=lm(bwt.oz~mht+parity+mpregwt+mage+as.factor(inc)+as.factor(mrace)+as.factor(smoke)+as.factor(med)+as.
factor(med)*as.factor(inc)+as.factor(smoke)*as.factor(mrace), data=smoking)
anova(wt_reg_5, wt_reg)
```

```
## Analysis of Variance Table
##
## Model 1: bwt.oz ~ mht + parity + mpregwt + mage + as.factor(inc) + as.factor(mrace) +
##     as.factor(smoke) + as.factor(med) + as.factor(med) * as.factor(inc) +
##     as.factor(smoke) * as.factor(mrace)
## Model 2: bwt.oz ~ mht + parity + log(mpregwt) + mage + as.factor(inc) +
##     as.factor(mrace) + as.factor(smoke) + as.factor(med) + as.factor(med) *
##     as.factor(inc) + as.factor(smoke) * as.factor(mrace) + parity *
##     mage
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    802 219781
## 2    801 219002  1    779.24 2.8501 0.09176 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The only interaction which can be considered as an important predictor is between income and education as it was excpected.

```
final_reg=lm(bwt.oz~mht+parity+mpregwt+mage+as.factor(inc)+as.factor(mrace)+as.factor(smoke)+as.factor(med)+a
s.factor(med)*as.factor(inc), data=smoking)
summary(final_reg)
```

```
##
## Call:
## lm(formula = bwt.oz ~ mht + parity + mpregwt + mage + as.factor(inc) +
##     as.factor(mrace) + as.factor(smoke) + as.factor(med) + as.factor(med) *
##     as.factor(inc), data = smoking)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -66.973  -9.451   0.000   9.616  50.883
##
## Coefficients: (16 not defined because of singularities)
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      24.20999   24.36913   0.993 0.320779
## mht                               1.07347    0.27740   3.870 0.000118 ***
## parity                            0.70278    0.41530   1.692 0.090990 .
## mpregwt                           0.10802    0.03359   3.216 0.001350 **
## mage                             -0.04472    0.13682  -0.327 0.743883
## as.factor(inc)1                  52.96336   32.01014   1.655 0.098399 .
## as.factor(inc)2                  12.14805   20.34901   0.597 0.550685
## as.factor(inc)3                  54.10766   24.14333   2.241 0.025291 *
## as.factor(inc)4                  -6.20718    8.46978  -0.733 0.463856
## as.factor(inc)5                 -24.58382   23.54247  -1.044 0.296691
## as.factor(inc)6                 -15.21785    8.64572  -1.760 0.078760 .
## as.factor(inc)7                  -7.60390    8.00578  -0.950 0.342498
## as.factor(inc)8                  -9.91333    9.79228  -1.012 0.311669
## as.factor(inc)9                 -20.26363    9.04527  -2.240 0.025347 *
## as.factor(mrace)6                 3.30290    3.60865   0.915 0.360322
## as.factor(mrace)7                -8.16173    1.62579  -5.020 6.36e-07 ***
## as.factor(mrace)8                -7.64759    3.16139  -2.419 0.015781 *
## as.factor(mrace)9                -2.32944    4.50277  -0.517 0.605064
## as.factor(smoke)1                -9.68104    1.19873  -8.076 2.43e-15 ***
## as.factor(med)1                  11.50184   18.08623   0.636 0.524993
## as.factor(med)2                  10.43042   18.03713   0.578 0.563240
## as.factor(med)3                  25.51563   23.58399   1.082 0.279619
## as.factor(med)4                  12.97202   17.81854   0.728 0.466821
## as.factor(med)5                  24.69607   18.29287   1.350 0.177383
## as.factor(med)7                 -46.85048   24.13909  -1.941 0.052624 .
## as.factor(inc)1:as.factor(med)1 -46.76493   32.74743  -1.428 0.153665
## as.factor(inc)2:as.factor(med)1 -10.67125   21.69484  -0.492 0.622938
## as.factor(inc)3:as.factor(med)1 -52.96044   25.45813  -2.080 0.037814 *
## as.factor(inc)4:as.factor(med)1  10.68279   11.62777   0.919 0.358511
## as.factor(inc)5:as.factor(med)1  33.64251   25.17960   1.336 0.181893
## as.factor(inc)6:as.factor(med)1  22.91791   12.93745   1.771 0.076865 .
## as.factor(inc)7:as.factor(med)1  38.06045   12.84429   2.963 0.003134 **
## as.factor(inc)8:as.factor(med)1 -18.30953   20.56166  -0.890 0.373480
## as.factor(inc)9:as.factor(med)1        NA         NA      NA       NA
## as.factor(inc)1:as.factor(med)2 -44.37692   32.71606  -1.356 0.175343
## as.factor(inc)2:as.factor(med)2  -1.10422   21.56015  -0.051 0.959166
## as.factor(inc)3:as.factor(med)2 -47.52500   25.17082  -1.888 0.059372 .
## as.factor(inc)4:as.factor(med)2  16.13354   11.18647   1.442 0.149624
## as.factor(inc)5:as.factor(med)2  29.55494   24.67328   1.198 0.231327
## as.factor(inc)6:as.factor(med)2  22.00631   11.81565   1.862 0.062900 .
## as.factor(inc)7:as.factor(med)2  12.69364   10.81518   1.174 0.240867
## as.factor(inc)8:as.factor(med)2  27.88089   14.03940   1.986 0.047382 *
## as.factor(inc)9:as.factor(med)2  33.17773   13.14133   2.525 0.011771 *
## as.factor(inc)1:as.factor(med)3 -65.13351   36.73234  -1.773 0.076574 .
## as.factor(inc)2:as.factor(med)3 -13.52611   27.21497  -0.497 0.619318
## as.factor(inc)3:as.factor(med)3 -74.09245   30.22310  -2.452 0.014437 *
## as.factor(inc)4:as.factor(med)3  -2.88103   19.88318  -0.145 0.884828
## as.factor(inc)5:as.factor(med)3  19.20520   29.63494   0.648 0.517131
## as.factor(inc)6:as.factor(med)3   4.56252   20.21240   0.226 0.821469
## as.factor(inc)7:as.factor(med)3  -2.62488   19.18417  -0.137 0.891203
## as.factor(inc)8:as.factor(med)3        NA         NA      NA       NA
## as.factor(inc)9:as.factor(med)3        NA         NA      NA       NA
## as.factor(inc)1:as.factor(med)4 -48.81608   32.79586  -1.488 0.137015
## as.factor(inc)2:as.factor(med)4  -4.26278   21.51980  -0.198 0.843028
```

```
## as.factor(inc)3:as.factor(med)4 -47.52955   25.04560  -1.898 0.058090 .
## as.factor(inc)4:as.factor(med)4   8.95752   10.97047   0.817 0.414449
## as.factor(inc)5:as.factor(med)4  29.40577   24.68464   1.191 0.233903
## as.factor(inc)6:as.factor(med)4  22.92984   11.46733   2.000 0.045881 *
## as.factor(inc)7:as.factor(med)4  10.73761   10.79948   0.994 0.320389
## as.factor(inc)8:as.factor(med)4  19.85665   15.15947   1.310 0.190619
## as.factor(inc)9:as.factor(med)4  33.55620   13.77917   2.435 0.015096 *
## as.factor(inc)1:as.factor(med)5 -58.09992   33.15474  -1.752 0.080088 .
## as.factor(inc)2:as.factor(med)5 -12.01453   22.08831  -0.544 0.586639
## as.factor(inc)3:as.factor(med)5 -61.35101   25.57599  -2.399 0.016676 *
## as.factor(inc)4:as.factor(med)5        NA         NA      NA       NA
## as.factor(inc)5:as.factor(med)5  20.35363   25.05207   0.812 0.416772
## as.factor(inc)6:as.factor(med)5        NA         NA      NA       NA
## as.factor(inc)7:as.factor(med)5        NA         NA      NA       NA
## as.factor(inc)8:as.factor(med)5        NA         NA      NA       NA
## as.factor(inc)9:as.factor(med)5        NA         NA      NA       NA
## as.factor(inc)1:as.factor(med)7        NA         NA      NA       NA
## as.factor(inc)2:as.factor(med)7  43.24555   31.79172   1.360 0.174122
## as.factor(inc)3:as.factor(med)7        NA         NA      NA       NA
## as.factor(inc)4:as.factor(med)7        NA         NA      NA       NA
## as.factor(inc)5:as.factor(med)7        NA         NA      NA       NA
## as.factor(inc)6:as.factor(med)7        NA         NA      NA       NA
## as.factor(inc)7:as.factor(med)7        NA         NA      NA       NA
## as.factor(inc)8:as.factor(med)7        NA         NA      NA       NA
## as.factor(inc)9:as.factor(med)7        NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.56 on 806 degrees of freedom
## Multiple R-squared:  0.2187, Adjusted R-squared:  0.1586
## F-statistic: 3.639 on 62 and 806 DF,  p-value: < 2.2e-16
```

```
#confint(final_reg)
```

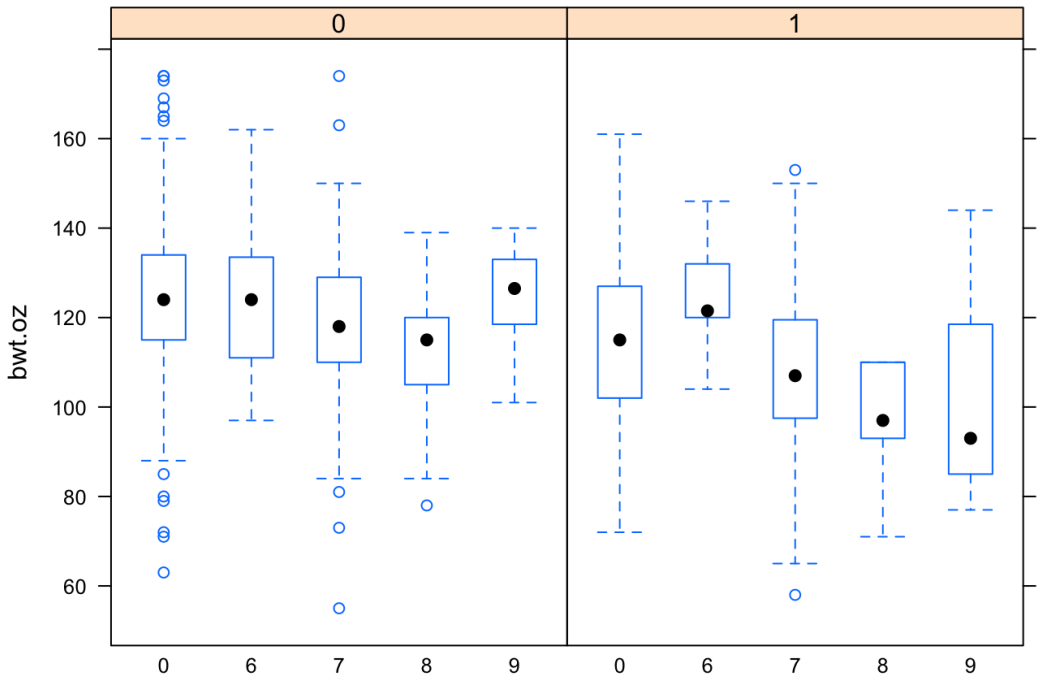
Based on the result above, the best fitting model is final_reg=lm(bwt.oz~mht+parity+mpregwt+mage+as.factor(inc)+as.factor(mrace)+as.factor(smoke)+as.factor(med)+as.factor(med)*as.factor(inc), data=smoking).

1. Do mothers who smoke tend to give birth to babies with lower weights than mothers who do not smoke? What is a likely range for the difference in birth weights for smokers and non-smokers?

Children of mothers who smoke on average weight 9.6 ounces less with 95% confidence interval (7.2, 11.2).

2. Is there any evidence that the association between smoking and birth weight differs by mother's race? If so, characterize those differences.

```
bwplot(bwt.oz ~ as.factor(mrace) | as.factor(smoke), data = smoking)
```




```
wt_reg_6=lm(bwt.oz~mht+parity+mpregwt+mage+as.factor(inc)+as.factor(mrace)+as.factor(smoke)+as.factor(med)+as.
factor(med)*as.factor(inc)+as.factor(smoke)*as.factor(mrace), data=smoking)
anova(wt_reg_6, final_reg)
```

```
## Analysis of Variance Table
##
## Model 1: bwt.oz ~ mht + parity + mpregwt + mage + as.factor(inc) + as.factor(mrace) +
##     as.factor(smoke) + as.factor(med) + as.factor(med) * as.factor(inc) +
##     as.factor(smoke) * as.factor(mrace)
## Model 2: bwt.oz ~ mht + parity + mpregwt + mage + as.factor(inc) + as.factor(mrace) +
##     as.factor(smoke) + as.factor(med) + as.factor(med) * as.factor(inc)
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    802 219781
## 2    806 220962 -4   -1180.8 1.0772 0.3666
```

Yes, there is an assocciation that smoking and birth weight differs by mother's race. Based on the boxplot, smoking effects children weights more if the mother has a mix race or she is an asian. Mother's smoking habit has the least effect on children whose mothers are mexicans. However, nested F-test does not support the hypothesis that children weights differ more if the mothers are smoking asian or black (probably due to the insufficient number of black and asian mothers), so we reject it.

3. Are there other interesting associations with birth weight that are worth mentioning? We suspected an interaction between mother's age and parity, however, nested F-test rejected that model. So the conclusion is that there is no significant associations affecting children weight.