Jianwen Wu
#83351
Prof. Foster

Homework 5 Nonlinear Regression.R

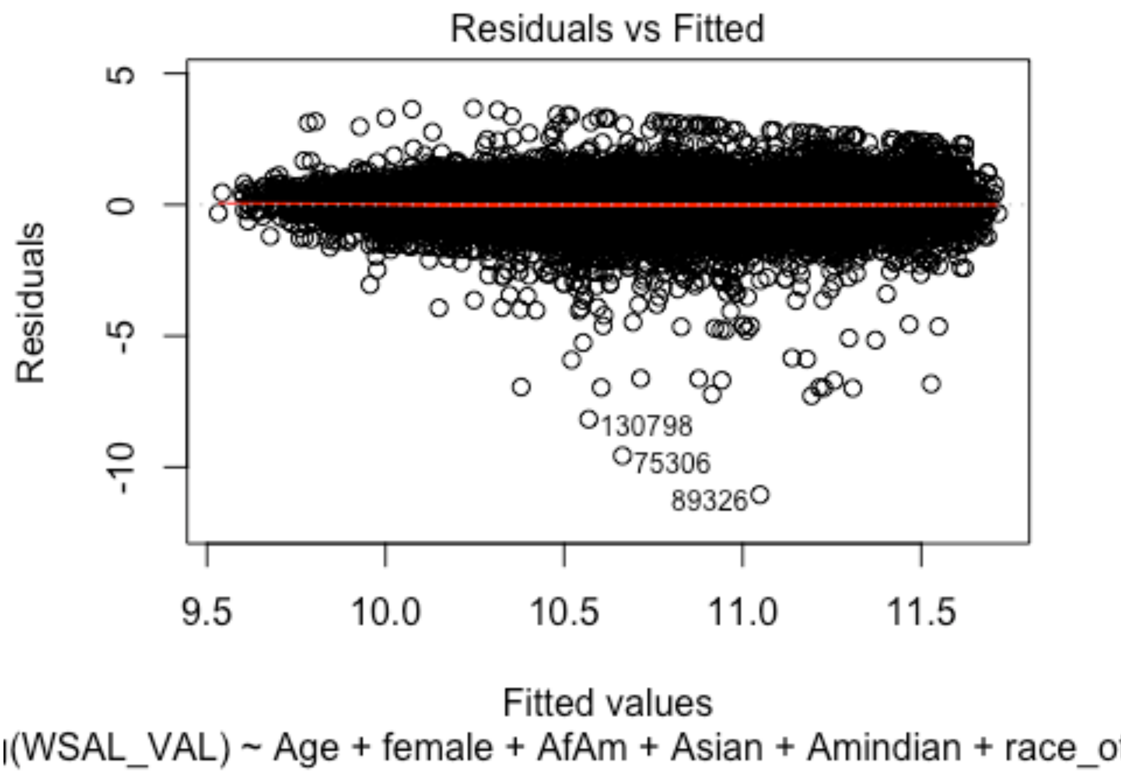
1. What are the names of the people currently in your study group? Jianwen Wu, Crystal Hernandez, Keely Allabt, Emmanuel Monroy.

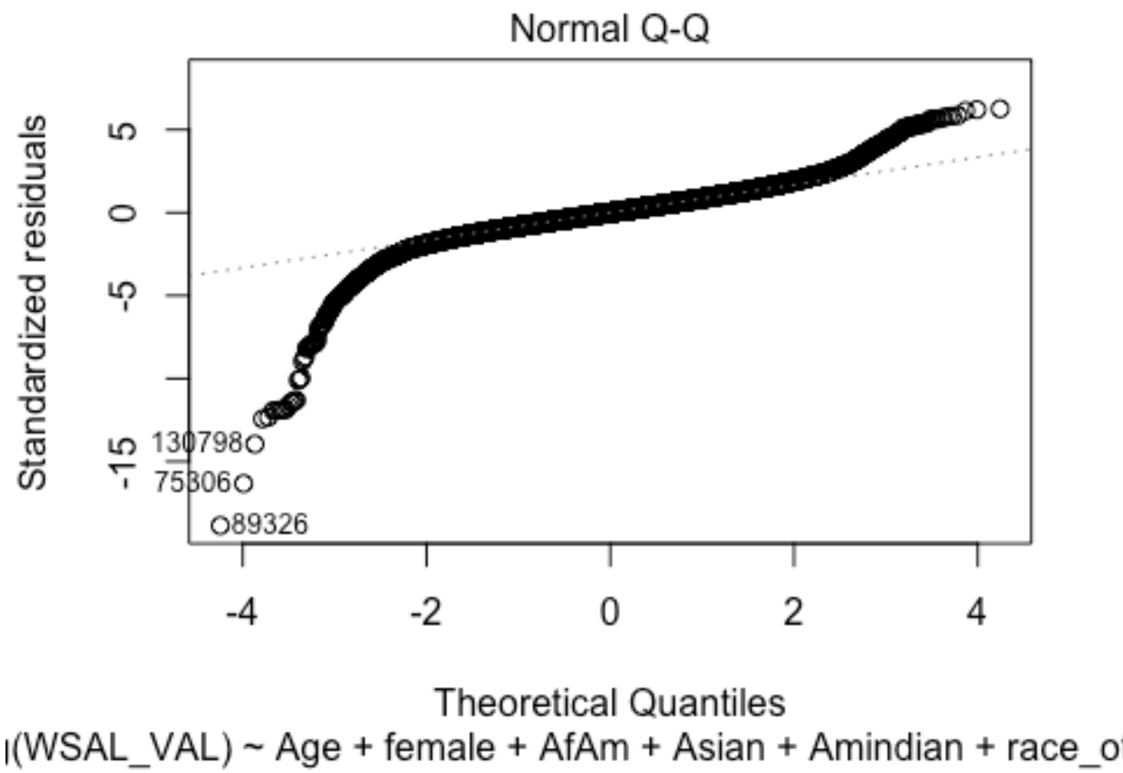
2. Using the CPS data, construct some interesting regressions on wage and salary (you might use the same subgroup as I did or you might change it up). Estimate a linear, quadratic, cubic and quartic specification of age on log wage. Don't just give me raw output! Make a nice table, like stargazer or in Stock and Watson (e.g. Chapter 9, Table 9.2). Make nice graphs and tests of groups of coefficients like I showed in class.

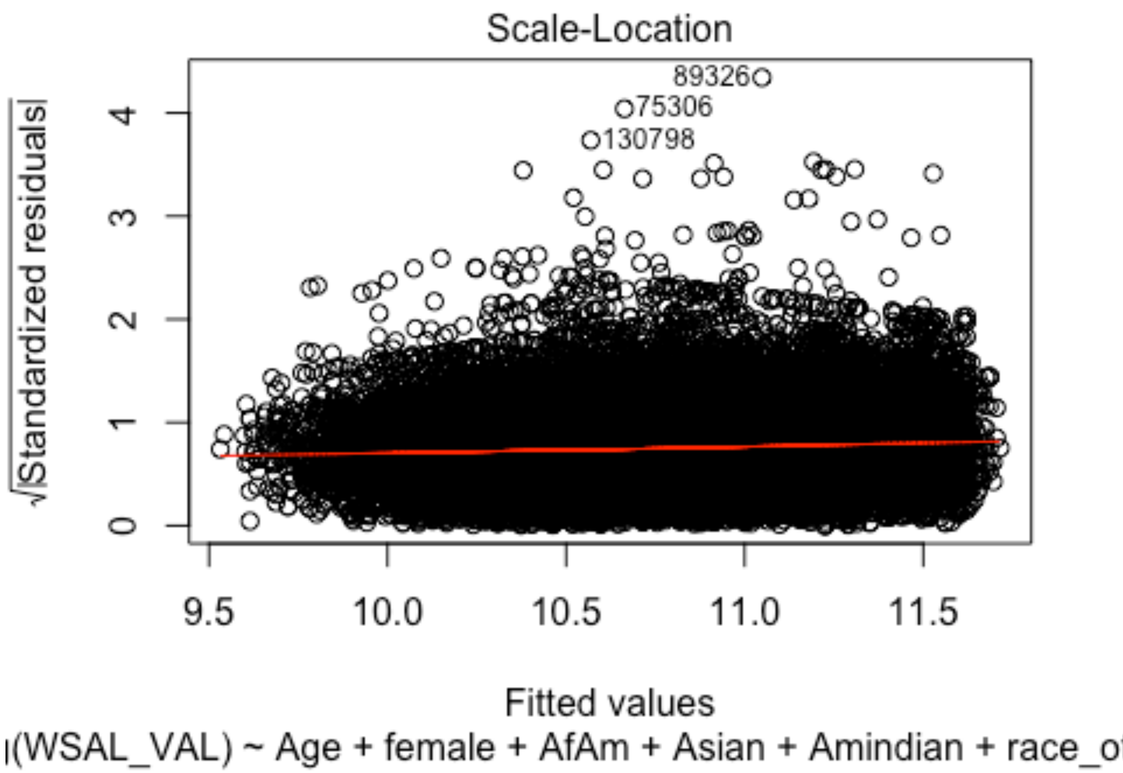
Excel with table annexed at Page 50.

```
load("/Users/jianwenwu/Desktop/ECO
B2000/cps_mar2013/cps_mar2013.RData")
attach(dat_CPSMar2013)
use_varb <- (Age >= 25) & (Age <= 55) & work_fullt & work_50wks
dat_use <- subset(dat_CPSMar2013,use_varb)
detach(dat_CPSMar2013)
attach(dat_use)
dat_noZeroWage <- subset(dat_use,(WSAL_VAL > 0))
detach(dat_use)
attach(dat_noZeroWage)

model1a <- lm(log(WSAL_VAL) ~ Age + female + AfAm + Asian + Amindian +
race_oth
              + Hispanic + educ_hs + educ_smcoll + educ_as + educ_bach
+ educ_adv

              + married + divwidsep + union_m + veteran + immigrant +
immig2gen, data = dat_noZeroWage)
summary(model1a)
##
## Call:
## lm(formula = log(WSAL_VAL) ~ Age + female + AfAm + Asian + Amindian
+
##     race_oth + Hispanic + educ_hs + educ_smcoll + educ_as + educ_bac
h +
##     educ_adv + married + divwidsep + union_m + veteran + immigrant +

##     immig2gen, data = dat_noZeroWage)
##
## Residuals:
```

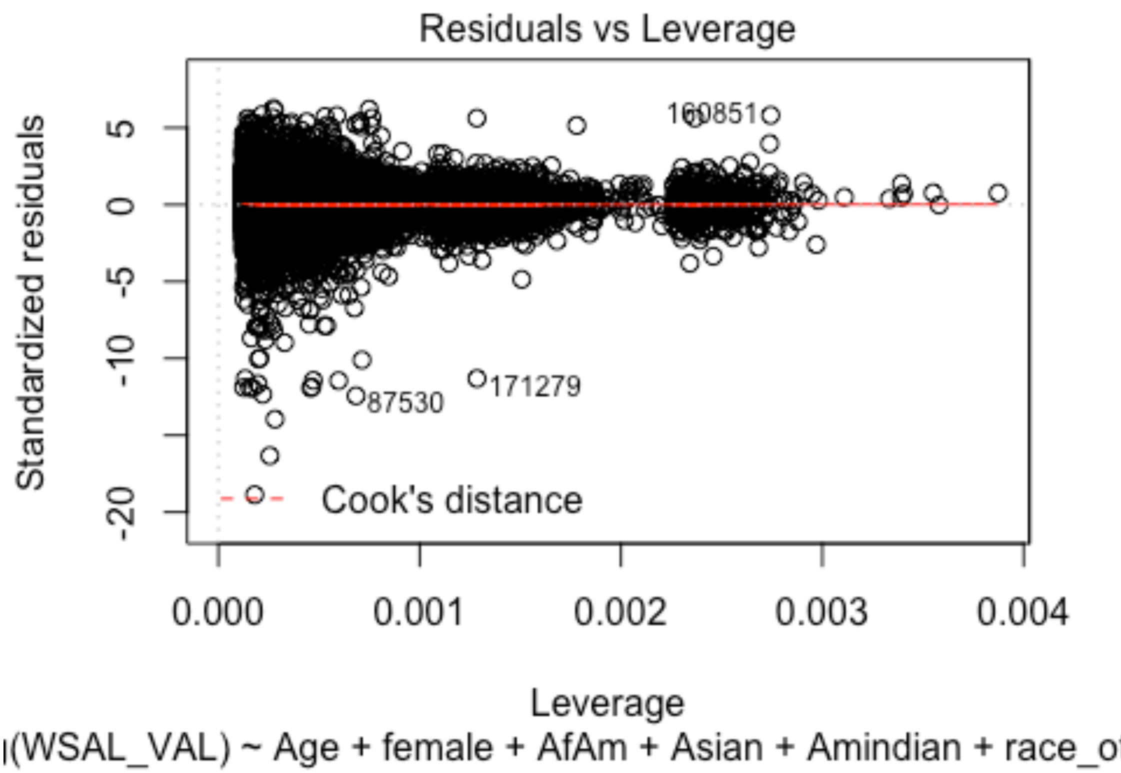
```
##      Min       1Q   Median       3Q      Max
## -11.0486  -0.3235   0.0029   0.3348   3.6645
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.8508146  0.0183002 538.291  < 2e-16 ***
## Age          0.0103909  0.0003413  30.442  < 2e-16 ***
## female      -0.3066790  0.0057144 -53.668  < 2e-16 ***
## AfAm        -0.1348315  0.0093219 -14.464  < 2e-16 ***
## Asian        0.0091999  0.0127948   0.719  0.47212
## Amindian    -0.0690166  0.0273319  -2.525  0.01157 *
## race_oth    -0.0521587  0.0187291  -2.785  0.00536 **
## Hispanic    -0.1183893  0.0094346 -12.548  < 2e-16 ***
## educ_hs      0.2552920  0.0126989  20.103  < 2e-16 ***
## educ_smcoll  0.4169565  0.0135501  30.772  < 2e-16 ***
## educ_as      0.4656438  0.0142824  32.603  < 2e-16 ***
## educ_bach    0.7622496  0.0129570  58.829  < 2e-16 ***
## educ_adv     1.0545368  0.0137985  76.424  < 2e-16 ***
## married      0.1445854  0.0074456  19.419  < 2e-16 ***
## divwidsep    0.0519897  0.0100821   5.157 2.52e-07 ***
## union_m      0.0307778  0.0184083   1.672  0.09454 .
## veteran      0.0297534  0.0119379   2.492  0.01269 *
## immigrant   -0.1456057  0.0122177 -11.918  < 2e-16 ***
## immig2gen    0.0747945  0.0114220   6.548 5.88e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5858 on 45517 degrees of freedom
## Multiple R-squared:  0.2977, Adjusted R-squared:  0.2974
## F-statistic:  1072 on 18 and 45517 DF,  p-value: < 2.2e-16

plot(model1a)
```


Residuals vs Fitted
Residuals
5
0
-5
-10
130798
75306
89326
9.5
10.0
10.5
11.0
11.5
Fitted values
(WSAL_VAL) ~ Age + female + AfAm + Asian + Amindian + race_o


Normal Q-Q
Standardized residuals
5
0
-5
-15
130798
75306
89326
-4
-2
0
2
4
Theoretical Quantiles
(WSAL_VAL) ~ Age + female + AfAm + Asian + Amindian + race_o




```r
require(car)

## Loading required package: car

## Warning: package 'car' was built under R version 3.2.5
```




```
require(ggplot2)

## Loading required package: ggplot2

model2a <- lm(log(WSAL_VAL) ~ Age+ I(Age^2) + female + AfAm + Asian + A
mindian + race_oth
              + Hispanic + educ_hs + educ_smcoll + educ_as + educ_bach
+ educ_adv

              + married + divwidsep + union_m + veteran + immigrant + i
mmig2gen, data = dat_noZeroWage)
summary(model2a)

##
## Call:
## lm(formula = log(WSAL_VAL) ~ Age + I(Age^2) + female + AfAm +
##     Asian + Amindian + race_oth + Hispanic + educ_hs + educ_smcoll +

##     educ_as + educ_bach + educ_adv + married + divwidsep + union_m +

##     veteran + immigrant + immig2gen, data = dat_noZeroWage)
##
## Residuals:
```

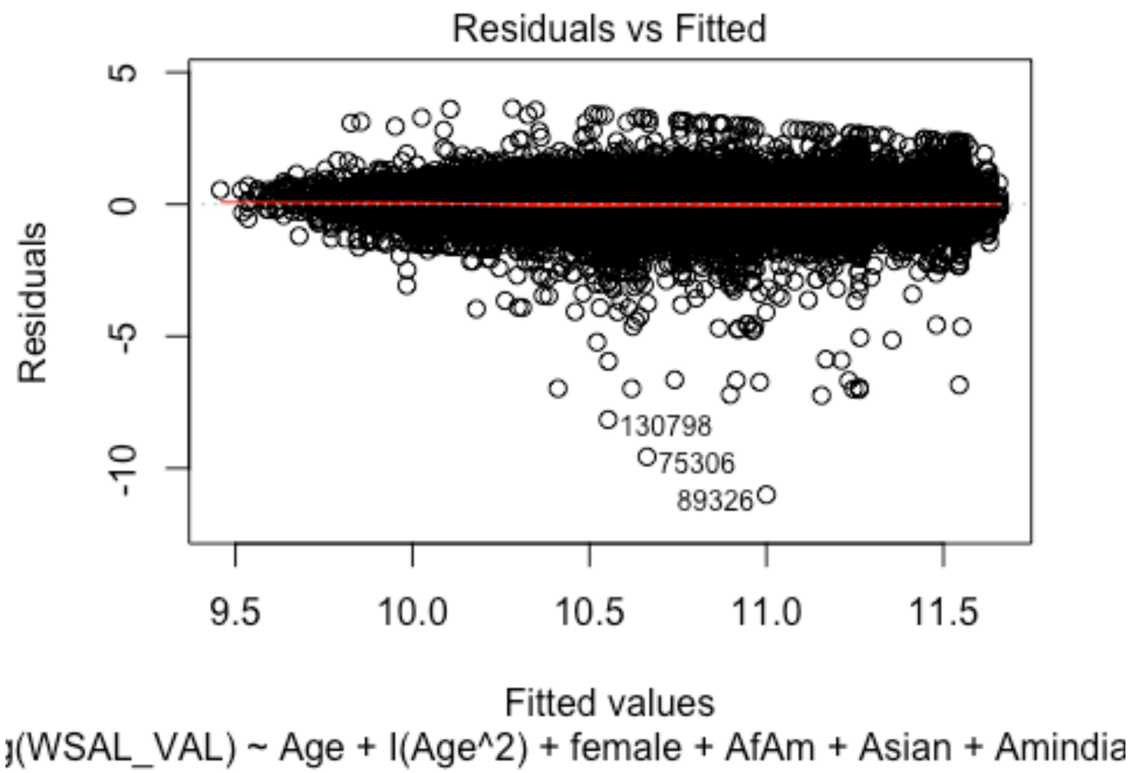
```
##      Min       1Q   Median       3Q      Max
## -10.9999  -0.3229   0.0012   0.3364   3.6283
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.015e+00  6.560e-02 137.411  < 2e-16 ***
## Age          5.440e-02  3.334e-03  16.317  < 2e-16 ***
## I(Age^2)    -5.421e-04  4.085e-05 -13.270  < 2e-16 ***
## female      -3.060e-01  5.704e-03 -53.641  < 2e-16 ***
## AfAm        -1.386e-01  9.308e-03 -14.892  < 2e-16 ***
## Asian        1.049e-02  1.277e-02   0.822 0.411349
## Amindian    -7.226e-02  2.728e-02  -2.649 0.008081 **
## race_oth    -5.246e-02  1.869e-02  -2.806 0.005011 **
## Hispanic    -1.195e-01  9.417e-03 -12.691  < 2e-16 ***
## educ_hs      2.559e-01  1.267e-02  20.193  < 2e-16 ***
## educ_smcoll  4.164e-01  1.352e-02  30.787  < 2e-16 ***
## educ_as      4.643e-01  1.426e-02  32.569  < 2e-16 ***
## educ_bach    7.613e-01  1.293e-02  58.866  < 2e-16 ***
## educ_adv     1.049e+00  1.378e-02  76.143  < 2e-16 ***
## married      1.258e-01  7.566e-03  16.622  < 2e-16 ***
## divwidsep    3.345e-02  1.016e-02   3.293 0.000992 ***
## union_m      3.154e-02  1.837e-02   1.716 0.086084 .
## veteran      2.959e-02  1.191e-02   2.484 0.013002 *
## immigrant   -1.513e-01  1.220e-02 -12.398  < 2e-16 ***
## immig2gen    7.656e-02  1.140e-02   6.715  1.9e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5846 on 45516 degrees of freedom
## Multiple R-squared:  0.3004, Adjusted R-squared:  0.3001
## F-statistic:  1029 on 19 and 45516 DF,  p-value: < 2.2e-16
```

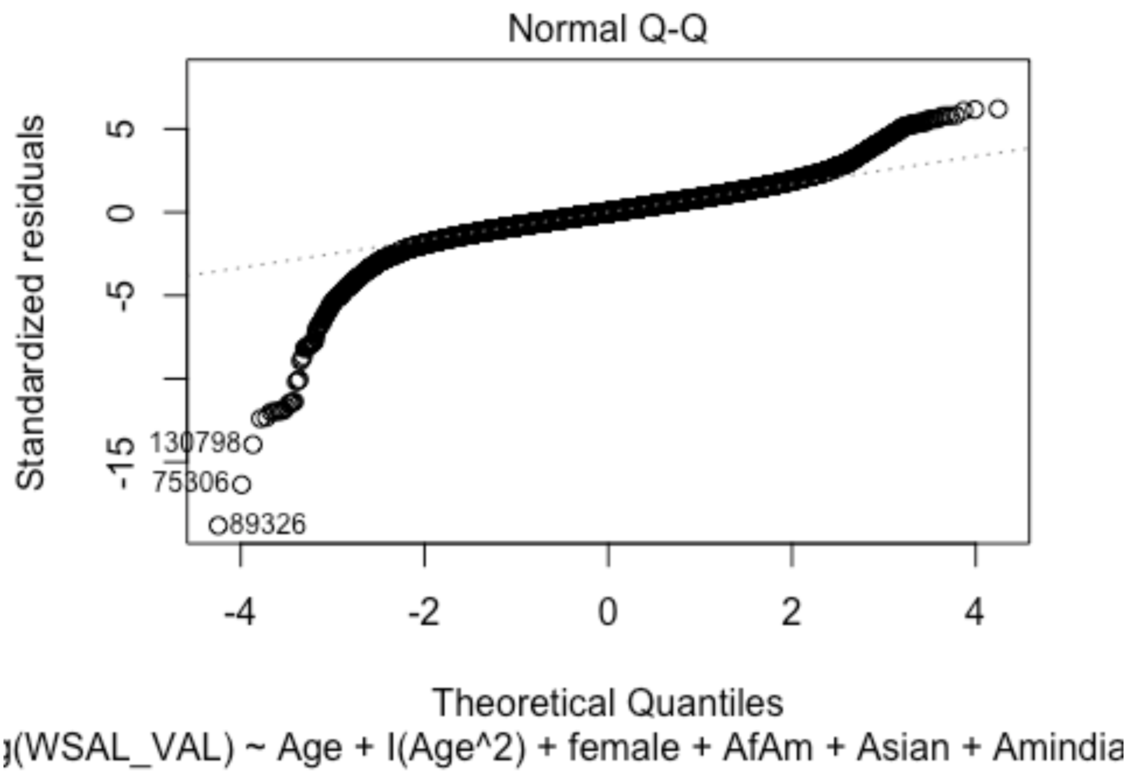
```
linearHypothesis(model2a, c('I(Age^2) = 0', ' Age = 0' ))
```

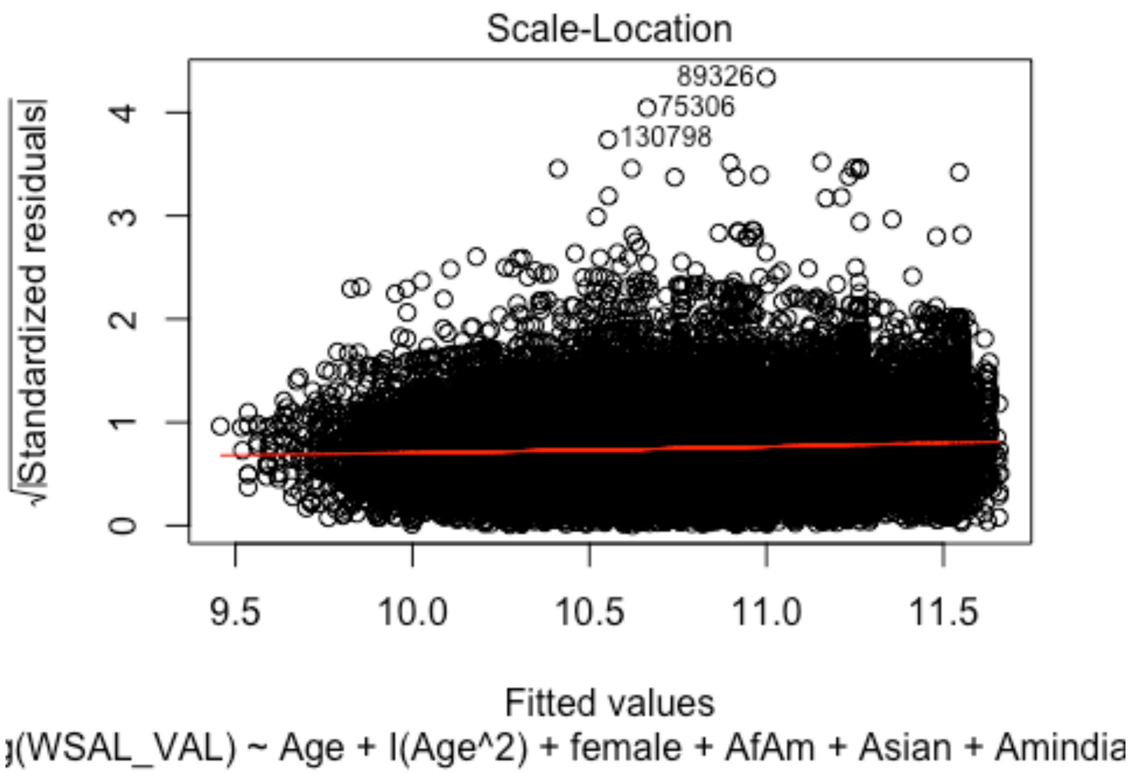
```
## Linear hypothesis test
##
## Hypothesis:
## I(Age^2) = 0
## Age = 0
##
## Model 1: restricted model
## Model 2: log(WSAL_VAL) ~ Age + I(Age^2) + female + AfAm + Asian + Am
indian +
##     race_oth + Hispanic + educ_hs + educ_smcoll + educ_as + educ_bac
h +
##     educ_adv + married + divwidsep + union_m + veteran + immigrant +

##     immig2gen
##
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
```

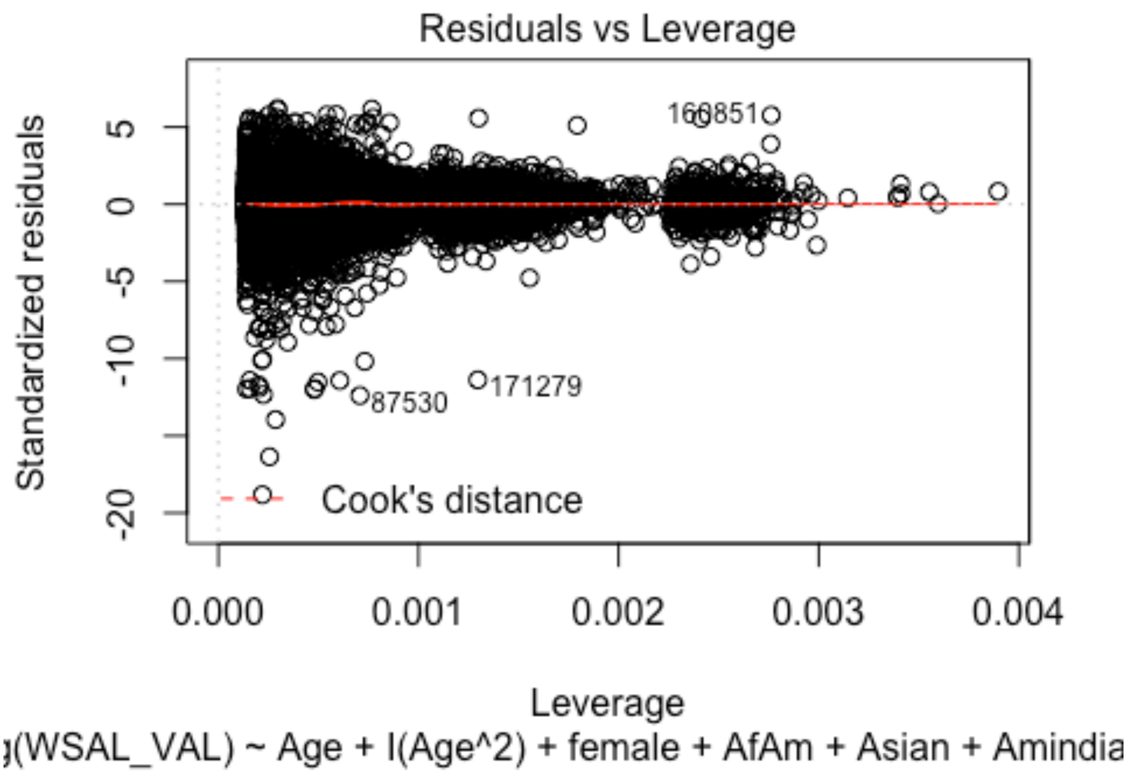
```
## 1  45518 15936
## 2  45516 15558  2    378.16 553.17 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plot(model2a)
```


Residuals vs Fitted
Residuals
5
0
-5
-10
130798
75306
89326
9.5
10.0
10.5
11.0
11.5
Fitted values
g(WSAL_VAL) ~ Age + I(Age^2) + female + AfAm + Asian + Amindia


Normal Q-Q
Standardized residuals
5
0
-5
-15
130798
75306
89326
-4
-2
0
2
4
Theoretical Quantiles
g(WSAL_VAL) ~ Age + I(Age^2) + female + AfAm + Asian + Amindia

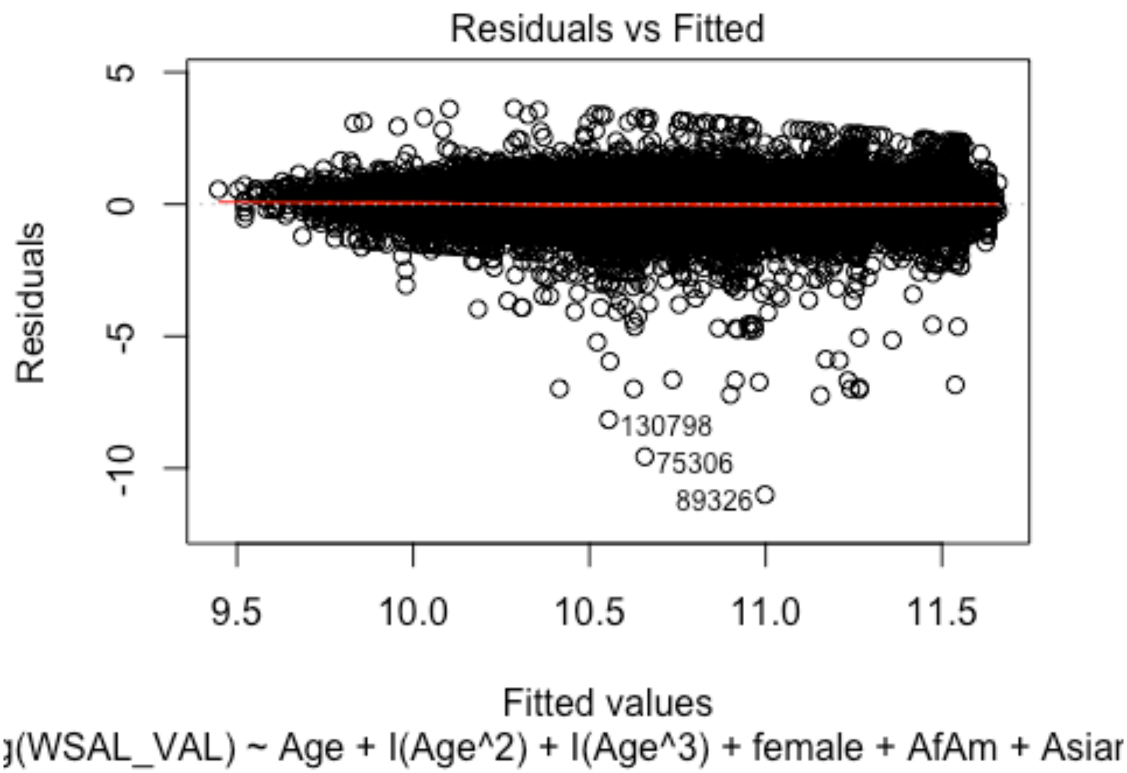

Scale-Location
89326
75306
130798
√|Standardized residuals|
0
1
2
3
4
9.5
10.0
10.5
11.0
11.5
Fitted values
g(WSAL_VAL) ~ Age + I(Age^2) + female + AfAm + Asian + Amindia

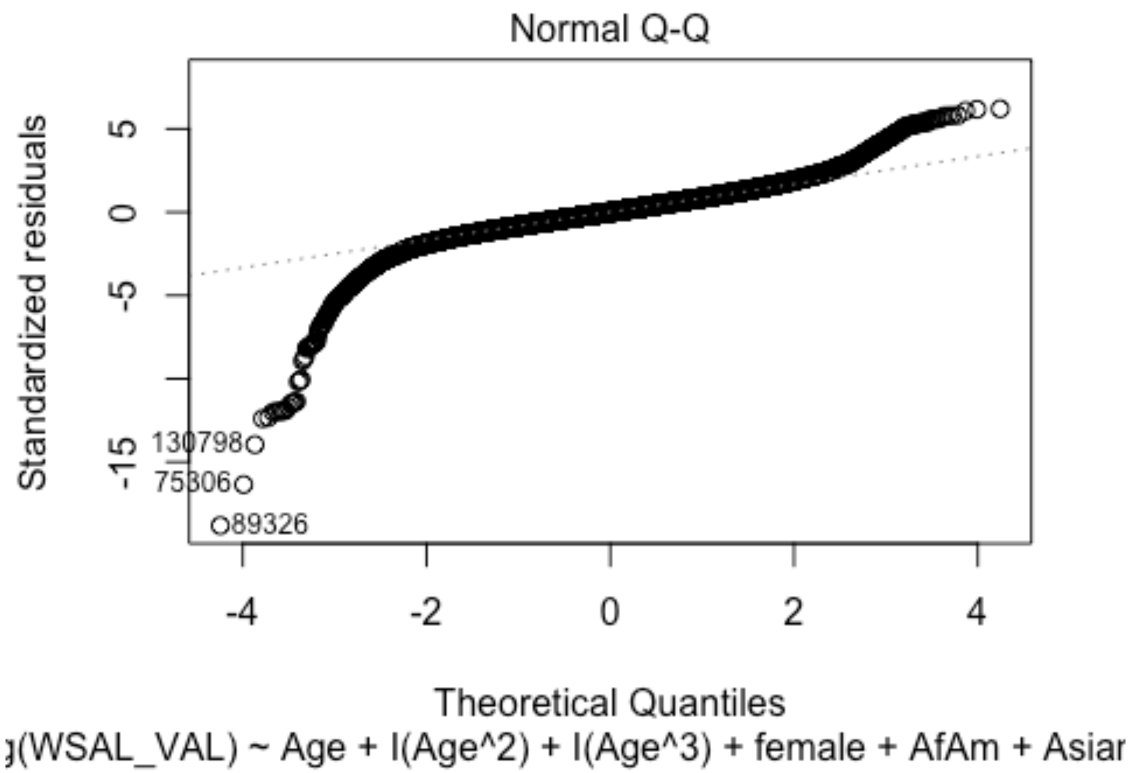



```r
model3a <- lm(log(WSAL_VAL) ~ Age + I(Age^2) + I(Age^3)
             + female
             + AfAm + Asian + Amindian + race_oth
             + Hispanic + educ_hs + educ_smcoll + educ_as + educ_bach +
 educ_adv
             + married + divwidsep + union_m + veteran + immigrant + im
mig2gen, data = dat_noZeroWage)
summary(model3a)
```

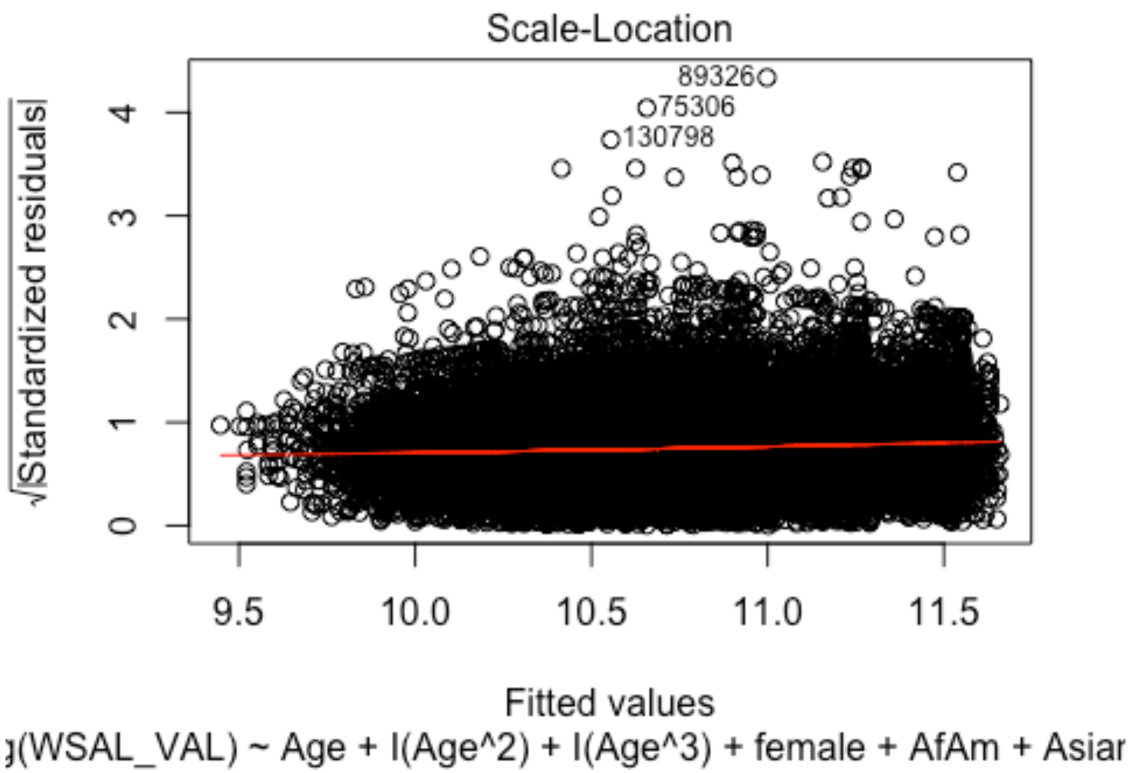
```
##
## Call:
## lm(formula = log(WSAL_VAL) ~ Age + I(Age^2) + I(Age^3) + female +
##     AfAm + Asian + Amindian + race_oth + Hispanic + educ_hs +
##     educ_smcoll + educ_as + educ_bach + educ_adv + married +
##     divwidsep + union_m + veteran + immigrant + immig2gen, data = da
t_noZeroWage)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.9981  -0.3230   0.0017   0.3356   3.6250
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
```

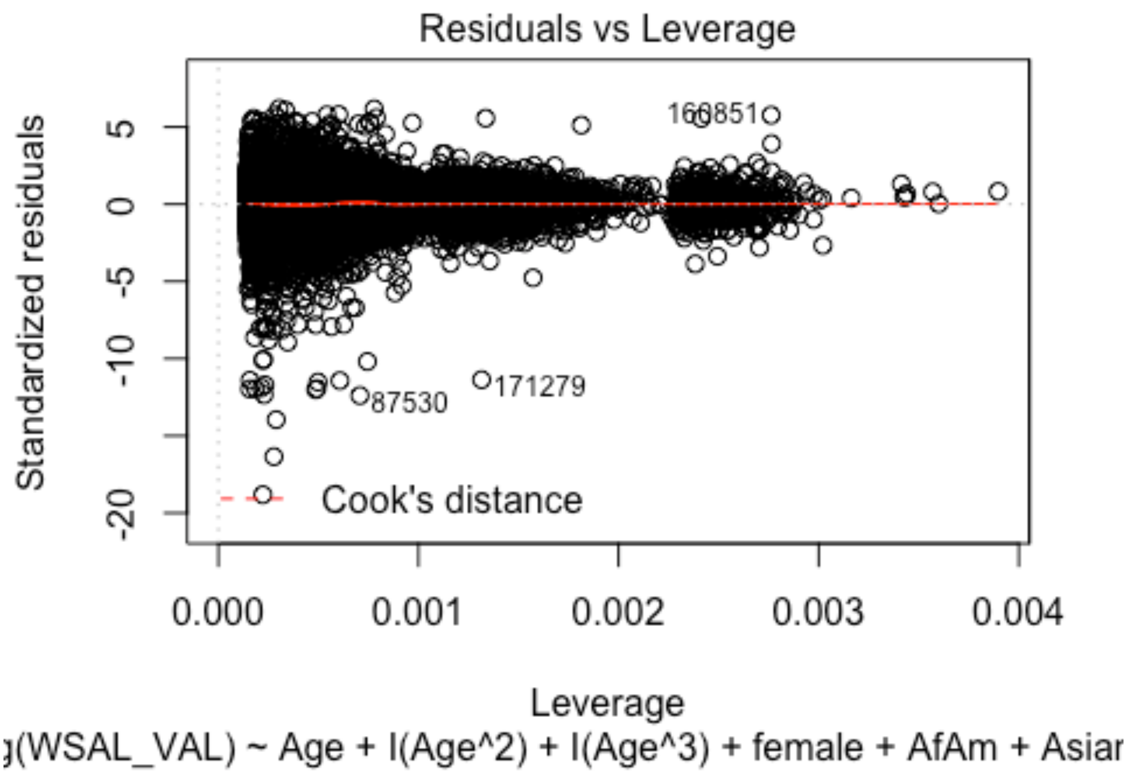
```
## (Intercept)  8.396e+00  3.116e-01  26.946  < 2e-16 ***
## Age           1.034e-01  2.434e-02   4.249 2.15e-05 ***
## I(Age^2)     -1.794e-03  6.171e-04  -2.907  0.00366 **
## I(Age^3)      1.036e-05  5.096e-06   2.033  0.04211 *
## female       -3.059e-01  5.704e-03 -53.640  < 2e-16 ***
## AfAm         -1.389e-01  9.309e-03 -14.916  < 2e-16 ***
## Asian         1.026e-02  1.277e-02   0.803  0.42184
## Amindian     -7.294e-02  2.728e-02  -2.674  0.00751 **
## race_oth     -5.271e-02  1.869e-02  -2.820  0.00481 **
## Hispanic     -1.194e-01  9.417e-03 -12.680  < 2e-16 ***
## educ_hs       2.562e-01  1.268e-02  20.216  < 2e-16 ***
## educ_smcoll   4.166e-01  1.352e-02  30.806  < 2e-16 ***
## educ_as       4.644e-01  1.426e-02  32.579  < 2e-16 ***
## educ_bach     7.616e-01  1.293e-02  58.890  < 2e-16 ***
## educ_adv      1.049e+00  1.378e-02  76.126  < 2e-16 ***
## married       1.246e-01  7.586e-03  16.427  < 2e-16 ***
## divwidsep     3.255e-02  1.017e-02   3.201  0.00137 **
## union_m       3.128e-02  1.837e-02   1.702  0.08870 .
## veteran       2.985e-02  1.192e-02   2.505  0.01225 *
## immigrant    -1.513e-01  1.220e-02 -12.397  < 2e-16 ***
## immig2gen     7.663e-02  1.140e-02   6.722 1.81e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5846 on 45515 degrees of freedom
## Multiple R-squared:  0.3005, Adjusted R-squared:  0.3002
## F-statistic: 977.5 on 20 and 45515 DF,  p-value: < 2.2e-16

plot(model3a)
```


Residuals vs Fitted
Residuals
5
0
-5
-10
130798
75306
89326
9.5
10.0
10.5
11.0
11.5
Fitted values
g(WSAL_VAL) ~ Age + I(Age^2) + I(Age^3) + female + AfAm + Asiar



Normal Q-Q
Standardized residuals
5
0
-5
-15
130798
75306
89326
-4
-2
0
2
4
Theoretical Quantiles
g(WSAL_VAL) ~ Age + I(Age^2) + I(Age^3) + female + AfAm + Asiar


Scale-Location
89326
75306
130798
√|Standardized residuals|
0
1
2
3
4
9.5
10.0
10.5
11.0
11.5
Fitted values
g(WSAL_VAL) ~ Age + I(Age^2) + I(Age^3) + female + AfAm + Asiar




```
linearHypothesis(model3a, c('I(Age^2) = 0', 'I(Age^3) = 0' ))

## Linear hypothesis test
##
## Hypothesis:
## I(Age^2) = 0
## I(Age^3) = 0
##
## Model 1: restricted model
## Model 2: log(WSAL_VAL) ~ Age + I(Age^2) + I(Age^3) + female + AfAm +
 Asian +
##     Amindian + race_oth + Hispanic + educ_hs + educ_smcoll +
##     educ_as + educ_bach + educ_adv + married + divwidsep + union_m +

##     veteran + immigrant + immig2gen
##
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1  45517 15618
## 2  45515 15556  2      61.6 90.115 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
coefs <- names(coef(model3a))
linearHypothesis(model3a, matchCoefs(model3a, "Age"), verbose=TRUE)

##
## Hypothesis matrix:
##            (Intercept) Age I(Age^2) I(Age^3) female AfAm Asian Amindia
n
## Age                  0   1        0        0      0    0     0
0
## I(Age^2)             0   0        1        0      0    0     0
0
## I(Age^3)             0   0        0        1      0    0     0
0
##            race_oth Hispanic educ_hs educ_smcoll educ_as educ_bach edu
c_adv
## Age               0        0       0           0       0         0
    0
## I(Age^2)          0        0       0           0       0         0
    0
## I(Age^3)          0        0       0           0       0         0
    0
##            married divwidsep union_m veteran immigrant immig2gen
## Age              0         0       0       0         0         0
## I(Age^2)         0         0       0       0         0         0
## I(Age^3)         0         0       0       0         0         0
##
## Right-hand-side vector:
## [1] 0 0 0
##
## Estimated linear function (hypothesis.matrix %*% coef - rhs)
##             Age      I(Age^2)      I(Age^3)
##  0.1033949194 -0.0017936435  0.0000103584
##
##
## Estimated variance/covariance matrix for linear function
##                      Age      I(Age^2)      I(Age^3)
## Age       5.922270e-04 -1.497908e-05  1.228534e-07
## I(Age^2) -1.497908e-05  3.808250e-07 -3.138094e-09
## I(Age^3)  1.228534e-07 -3.138094e-09  2.597250e-11

## Linear hypothesis test
##
## Hypothesis:
## Age = 0
## I(Age^2) = 0
## I(Age^3) = 0
##
## Model 1: restricted model
## Model 2: log(WSAL_VAL) ~ Age + I(Age^2) + I(Age^3) + female + AfAm +
 Asian +
```

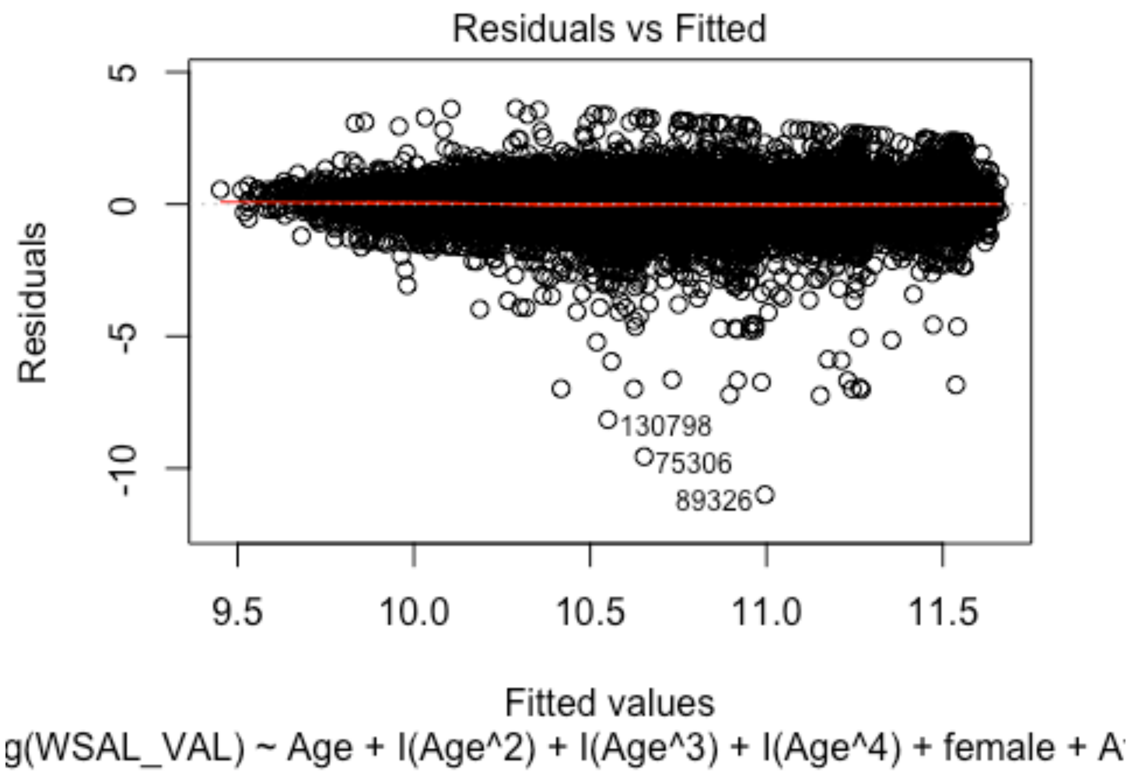
```
##     Amindian + race_oth + Hispanic + educ_hs + educ_smcoll +
##     educ_as + educ_bach + educ_adv + married + divwidsep + union_m +

##     veteran + immigrant + immig2gen
##
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1  45518 15936
## 2  45515 15556  3    379.57 370.18 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model4a <- lm(log(WSAL_VAL) ~ Age + I(Age^2) + I(Age^3) + I(Age^4)
              + female
              + AfAm + Asian + Amindian + race_oth
              + Hispanic + educ_hs + educ_smcoll + educ_as + educ_bach
+ educ_adv
              + married + divwidsep + union_m + veteran + immigrant + i
mmig2gen, data = dat_noZeroWage)
summary(model4a)
```

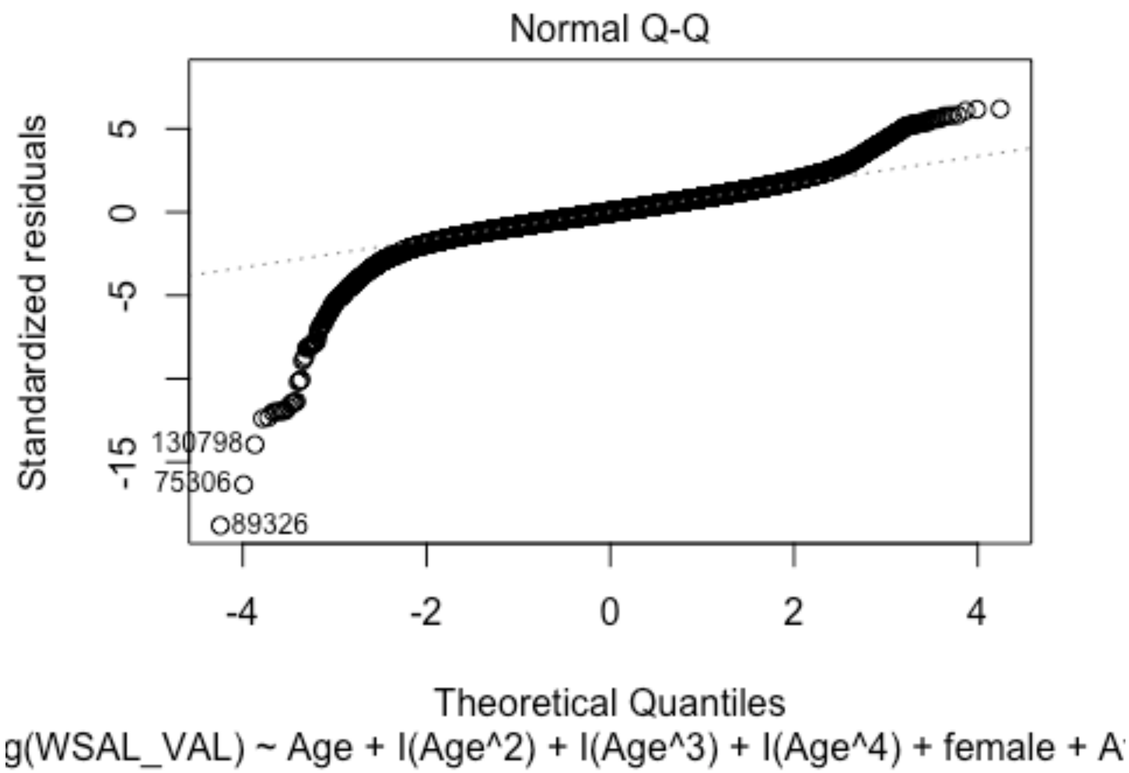
```
##
## Call:
## lm(formula = log(WSAL_VAL) ~ Age + I(Age^2) + I(Age^3) + I(Age^4) +
##     female + AfAm + Asian + Amindian + race_oth + Hispanic +
##     educ_hs + educ_smcoll + educ_as + educ_bach + educ_adv +
##     married + divwidsep + union_m + veteran + immigrant + immig2gen,

##     data = dat_noZeroWage)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.9953  -0.3221   0.0010   0.3360   3.6218
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.039e+01  1.517e+00   6.846 7.69e-12 ***
## Age         -1.088e-01  1.600e-01  -0.680  0.49643
## I(Age^2)     6.484e-03  6.199e-03   1.046  0.29558
## I(Age^3)    -1.300e-04  1.047e-04  -1.241  0.21445
## I(Age^4)     8.741e-07  6.513e-07   1.342  0.17962
## female      -3.059e-01  5.704e-03 -53.635  < 2e-16 ***
## AfAm        -1.390e-01  9.309e-03 -14.928  < 2e-16 ***
## Asian        1.025e-02  1.277e-02   0.803  0.42216
## Amindian    -7.276e-02  2.728e-02  -2.667  0.00765 **
## race_oth    -5.270e-02  1.869e-02  -2.819  0.00482 **
## Hispanic    -1.194e-01  9.417e-03 -12.677  < 2e-16 ***
## educ_hs      2.562e-01  1.268e-02  20.214  < 2e-16 ***
## educ_smcoll  4.166e-01  1.352e-02  30.801  < 2e-16 ***
## educ_as      4.645e-01  1.425e-02  32.582  < 2e-16 ***
## educ_bach    7.617e-01  1.293e-02  58.892  < 2e-16 ***
```

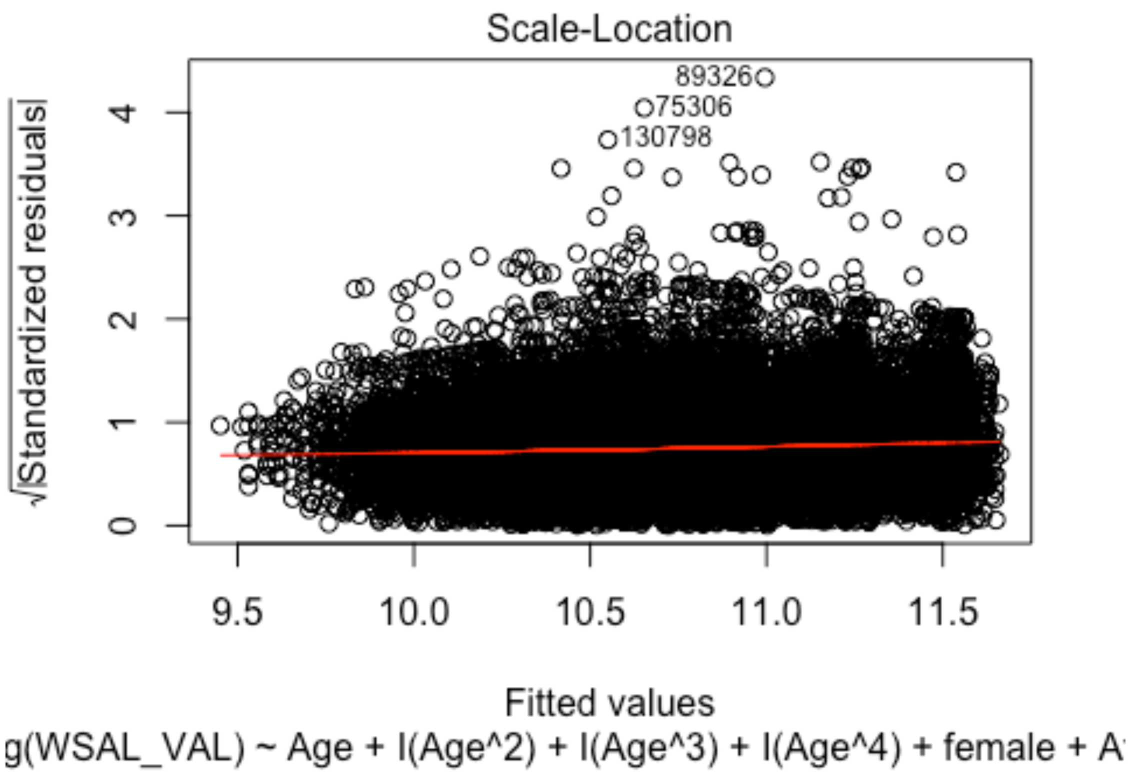
```
## educ_adv     1.049e+00  1.378e-02  76.131  < 2e-16 ***
## married      1.248e-01  7.587e-03  16.446  < 2e-16 ***
## divwidsep    3.268e-02  1.017e-02   3.214  0.00131 **
## union_m      3.115e-02  1.837e-02   1.696  0.08996 .
## veteran      2.989e-02  1.192e-02   2.508  0.01213 *
## immigrant   -1.513e-01  1.220e-02 -12.399  < 2e-16 ***
## immig2gen    7.665e-02  1.140e-02   6.723 1.80e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5846 on 45514 degrees of freedom
## Multiple R-squared:  0.3005, Adjusted R-squared:  0.3002
## F-statistic:   931 on 21 and 45514 DF,  p-value: < 2.2e-16
```

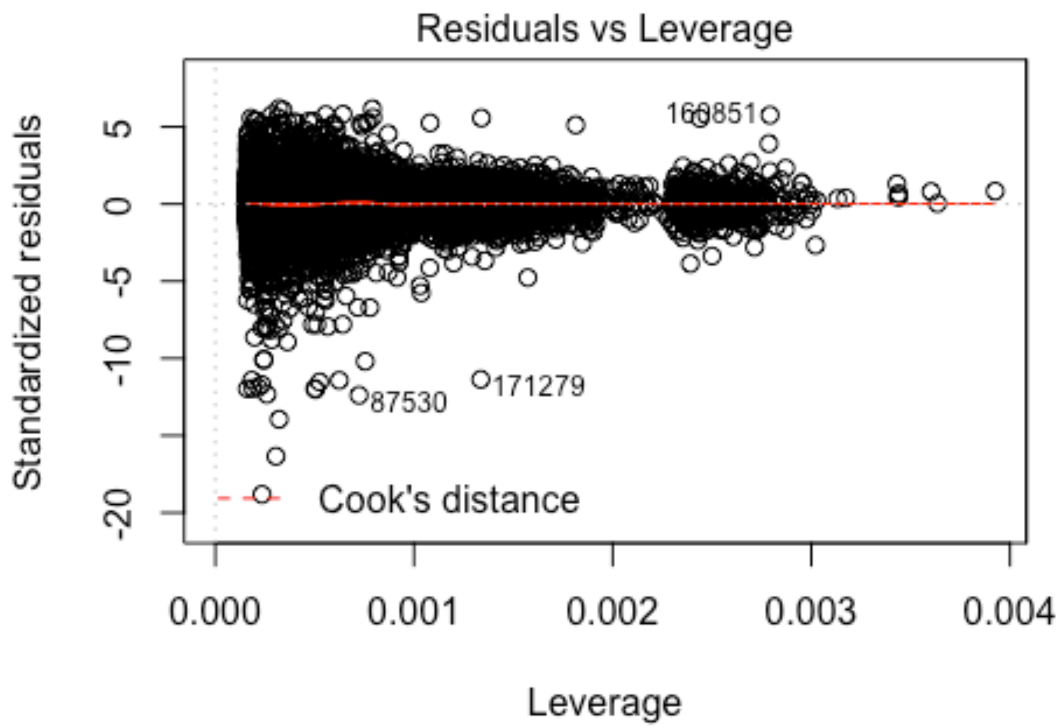
```r
plot(model4a)
```



Residuals vs Fitted
Residuals
5
0
-5
-10
9.5
10.0
10.5
11.0
11.5
130798
75306
89326
Fitted values
g(WSAL_VAL) ~ Age + I(Age^2) + I(Age^3) + I(Age^4) + female + A


Normal Q-Q
Standardized residuals
5
0
-5
-15
130798
75306
89326
-4
-2
0
2
4
Theoretical Quantiles
g(WSAL_VAL) ~ Age + I(Age^2) + I(Age^3) + I(Age^4) + female + A

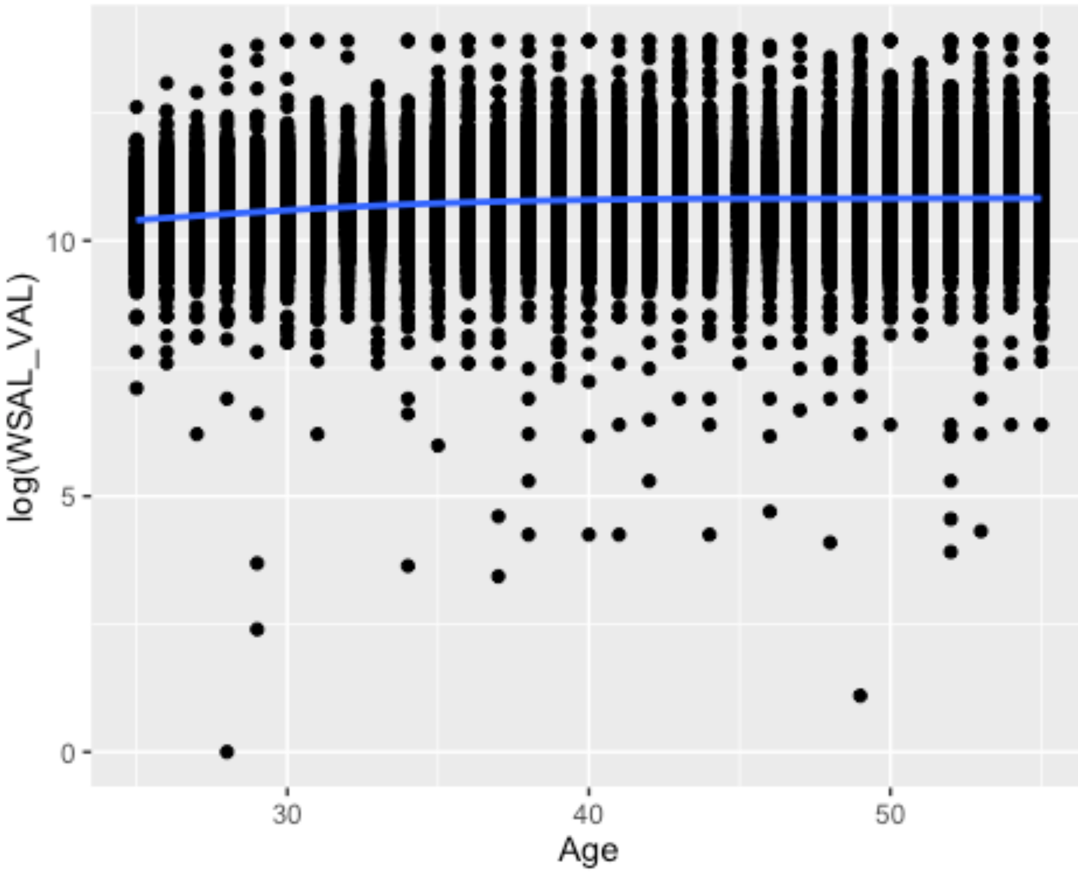

Scale-Location
89326
75306
130798
√|Standardized residuals|
0
1
2
3
4
9.5
10.0
10.5
11.0
11.5
Fitted values
g(WSAL_VAL) ~ Age + I(Age^2) + I(Age^3) + I(Age^4) + female + A



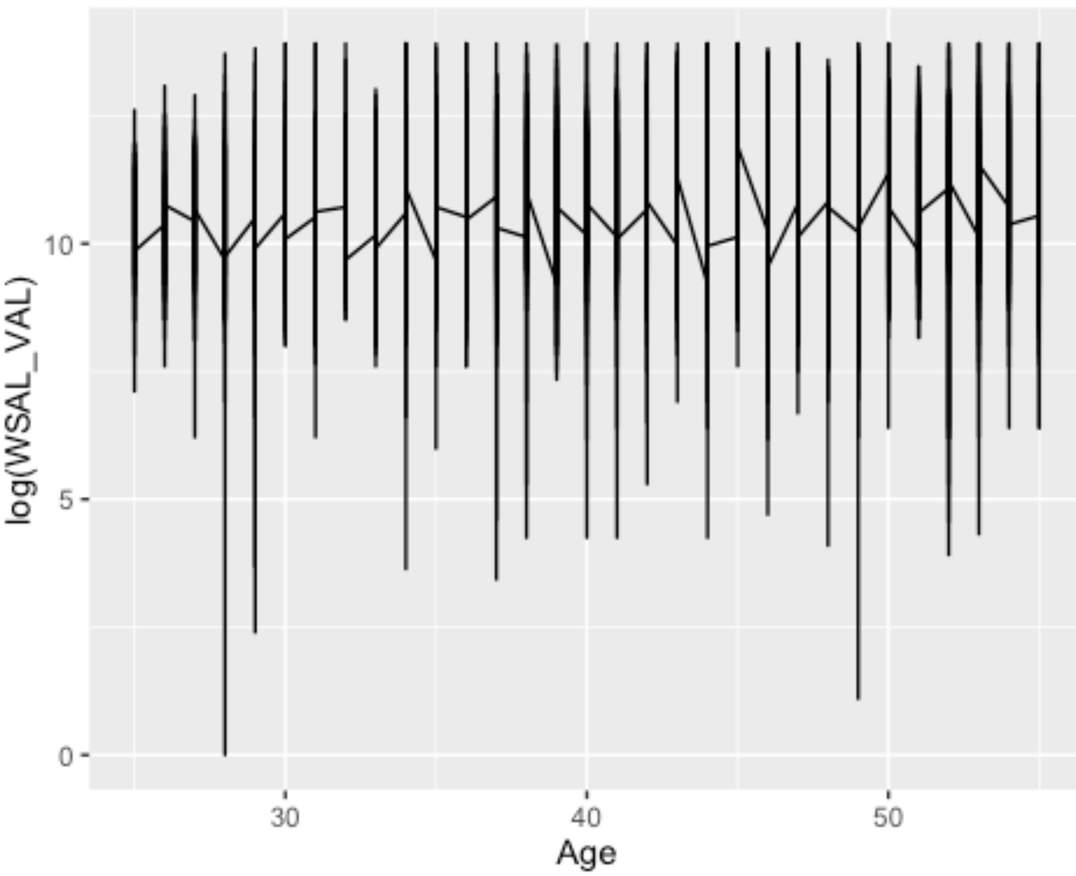
g(WSAL_VAL) ~ Age + I(Age^2) + I(Age^3) + I(Age^4) + female + A

```
linearHypothesis(model4a, c('I(Age^2) = 0', 'I(Age^3) = 0', 'I(Age^4) =
 0' ))

## Linear hypothesis test
##
## Hypothesis:
## I(Age^2) = 0
## I(Age^3) = 0
## I(Age^4) = 0
##
## Model 1: restricted model
## Model 2: log(WSAL_VAL) ~ Age + I(Age^2) + I(Age^3) + I(Age^4) + fema
le +
##     AfAm + Asian + Amindian + race_oth + Hispanic + educ_hs +
##     educ_smcoll + educ_as + educ_bach + educ_adv + married +
##     divwidsep + union_m + veteran + immigrant + immig2gen
##
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1  45517 15618
## 2  45514 15556  3    62.216 60.678 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
qplot(Age, log(WSAL_VAL), data = dat_noZeroWage, geom = c("point", "smo
oth"))
```




```
qplot(Age, log(WSAL_VAL), data = dat_noZeroWage, geom = "line")
```


log(WSAL_VAL)
10
5
0
30
40
50
Age

(3) **Next some interactions – your choice, try some interesting ones – you have data on gender, race, ethnicity, education, marital status, union, veteran, immigrant, detailed industry and occupation, state [GESTCEN], even metro area [GTCBSA]. Show me some nice output.**

Excel with table annexed at Page 50.

Nonlinear regression is in essence a type of regression analysis in which the data is modeled by a function which is a nonlinear combination and is dependent on one or more independent variables. So, the data is "fitted" with successive approximations, as evidenced by our various models. This is apparent in our model2 (shown below) where we put a nonlinear model into a linear regression. We decided to add the following interactions into the mix: gender, race, ethnicity, education, marital status, union, veteran, immigration status. We then use the ANOVA function to compare nested models. Our results are below:

```
require(AER)

## Loading required package: AER

## Loading required package: car

## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

## Loading required package: survival

load("cps_mar2013.RData")

attach(dat_CPSMar2013)

## The following object is masked from package:survival:
##
##     veteran


use_varb <- (Age >= 25) & (Age <= 55) & work_fullt & work_50wks
dat_use <- subset(dat_CPSMar2013,use_varb)

detach(dat_CPSMar2013)
```

```r
attach(dat_use)

## The following object is masked from package:survival:
##
##     veteran


summary(WSAL_VAL)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0   28020   45000   57310   70000 1100000

summary(Age)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   25.00   34.00   41.00   40.78   48.00   55.00

summary(female)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.4273  1.0000  1.0000

summary(AfAm)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.1029  0.0000  1.0000

summary(Asian)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.0718  0.0000  1.0000

summary(Amindian)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.00000 0.00000 0.01022 0.00000 1.00000

summary(race_oth)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.00000 0.00000 0.02225 0.00000 1.00000

summary(Hispanic)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.1655  0.0000  1.0000

summary(educ_hs)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.2521  1.0000  1.0000

summary(educ_smcoll)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##  0.0000  0.0000  0.0000  0.1604  0.0000  1.0000

summary(educ_as)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##  0.0000  0.0000  0.0000  0.1152  0.0000  1.0000

summary(educ_bach)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##  0.0000  0.0000  0.0000  0.2598  1.0000  1.0000

summary(educ_adv)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##  0.0000  0.0000  0.0000  0.1447  0.0000  1.0000

summary(married)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##  0.0000  0.0000  1.0000  0.6615  1.0000  1.0000

summary(divwidsep)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##  0.0000  0.0000  0.0000  0.1371  0.0000  1.0000

summary(union_m)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
## 0.00000 0.00000 0.00000 0.02185 0.00000 1.00000

summary(veteran)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
## 0.00000 0.00000 0.00000 0.05893 0.00000 1.00000

summary(immigrant)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##  0.0000  0.0000  0.0000  0.1875  0.0000  1.0000

summary(immig2gen)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##  0.0000  0.0000  0.0000  0.2509  1.0000  1.0000

model1 <- lm(WSAL_VAL ~ Age + female + AfAm + Asian + Amindian + race_o
th
             + Hispanic + educ_hs + educ_smcoll + educ_as + educ_bach +
 educ_adv
             + married + divwidsep + union_m + veteran + immigrant + im
mig2gen)
```

```r
summary(model1)
```

```
##
## Call:
## lm(formula = WSAL_VAL ~ Age + female + AfAm + Asian + Amindian +
##     race_oth + Hispanic + educ_hs + educ_smcoll + educ_as + educ_bac
h +
##     educ_adv + married + divwidsep + union_m + veteran + immigrant +

##     immig2gen)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -123822  -22631   -6241   11283 1075277
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6656.73    1801.46   3.695  0.00022 ***
## Age            711.93      33.73  21.108  < 2e-16 ***
## female      -19334.07     565.14 -34.211  < 2e-16 ***
## AfAm         -8679.40     925.52  -9.378  < 2e-16 ***
## Asian        -1625.51    1264.07  -1.286  0.19847
## Amindian       329.19    2704.50   0.122  0.90312
## race_oth     -4032.07    1855.04  -2.174  0.02974 *
## Hispanic     -6719.13     933.50  -7.198 6.21e-13 ***
## educ_hs       9117.64    1241.65   7.343 2.12e-13 ***
## educ_smcoll  17993.55    1327.20  13.557  < 2e-16 ***
## educ_as      19046.51    1400.74  13.597  < 2e-16 ***
## educ_bach    39285.83    1269.56  30.944  < 2e-16 ***
## educ_adv     67575.27    1355.90  49.838  < 2e-16 ***
## married       6767.08     738.77   9.160  < 2e-16 ***
## divwidsep     1228.63     997.57   1.232  0.21809
## union_m      -2575.03    1857.28  -1.386  0.16562
## veteran        182.37    1180.67   0.154  0.87724
## immigrant    -6094.72    1209.60  -5.039 4.71e-07 ***
## immig2gen     5090.24    1133.14   4.492 7.07e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 59130 on 47531 degrees of freedom
## Multiple R-squared:  0.1474, Adjusted R-squared:  0.1471
## F-statistic: 456.5 on 18 and 47531 DF,  p-value: < 2.2e-16
```

```r
coeftest(model1)
```

```
##
## t test of coefficients:
##
##                 Estimate Std. Error  t value  Pr(>|t|)
```

```
## (Intercept)   6656.733   1801.455   3.6952   0.00022 ***
## Age            711.935     33.728  21.1082 < 2.2e-16 ***
## female      -19334.071    565.141 -34.2111 < 2.2e-16 ***
## AfAm         -8679.399    925.517  -9.3779 < 2.2e-16 ***
## Asian        -1625.510   1264.071  -1.2859   0.19847
## Amindian       329.194   2704.497   0.1217   0.90312
## race_oth     -4032.066   1855.040  -2.1736   0.02974 *
## Hispanic     -6719.129    933.497  -7.1978 6.209e-13 ***
## educ_hs       9117.636   1241.647   7.3432 2.119e-13 ***
## educ_smcoll  17993.547   1327.204  13.5575 < 2.2e-16 ***
## educ_as      19046.514   1400.739  13.5975 < 2.2e-16 ***
## educ_bach    39285.835   1269.561  30.9444 < 2.2e-16 ***
## educ_adv     67575.269   1355.902  49.8379 < 2.2e-16 ***
## married       6767.082    738.767   9.1600 < 2.2e-16 ***
## divwidsep     1228.629    997.565   1.2316   0.21809
## union_m      -2575.027   1857.285  -1.3864   0.16562
## veteran        182.371   1180.673   0.1545   0.87724
## immigrant    -6094.717   1209.605  -5.0386 4.707e-07 ***
## immig2gen     5090.245   1133.141   4.4922 7.067e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


dat_noZeroWage <- subset(dat_use,(WSAL_VAL > 0))
model1a <- lm(log(WSAL_VAL) ~ Age + female + AfAm + Asian + Amindian +
race_oth
              + Hispanic + educ_hs + educ_smcoll + educ_as + educ_bach
+ educ_adv
              + married + divwidsep + union_m + veteran + immigrant + i
mmig2gen, data = dat_noZeroWage)
detach(dat_use)
attach(dat_noZeroWage)

## The following object is masked from package:survival:
##
##     veteran

log(mean(WSAL_VAL))

## [1] 10.99956

mean(log(WSAL_VAL))

## [1] 10.74171

detach(dat_noZeroWage)
attach(dat_use)

## The following object is masked from package:survival:
##
##     veteran
```

```r
require(sandwich)
require(lmtest)

coeftest(model1,vcovHC)

##
## t test of coefficients:
##
##                Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)   6656.733   1433.365   4.6441 3.424e-06 ***
## Age            711.935     31.137  22.8649 < 2.2e-16 ***
## female      -19334.071    539.294 -35.8507 < 2.2e-16 ***
## AfAm         -8679.399    646.035 -13.4349 < 2.2e-16 ***
## Asian        -1625.510   1310.443  -1.2404  0.214823
## Amindian       329.194   3471.314   0.0948  0.924448
## race_oth     -4032.066   1490.067  -2.7060  0.006813 **
## Hispanic     -6719.129    933.994  -7.1940 6.386e-13 ***
## educ_hs       9117.636    787.394  11.5795 < 2.2e-16 ***
## educ_smcoll  17993.547    915.721  19.6496 < 2.2e-16 ***
## educ_as      19046.514    915.212  20.8110 < 2.2e-16 ***
## educ_bach    39285.835    945.628  41.5447 < 2.2e-16 ***
## educ_adv     67575.269   1468.396  46.0198 < 2.2e-16 ***
## married       6767.082    603.297  11.2168 < 2.2e-16 ***
## divwidsep     1228.629    838.934   1.4645  0.143061
## union_m      -2575.027   1541.328  -1.6707  0.094796 .
## veteran        182.371   1313.661   0.1388  0.889588
## immigrant    -6094.717   1263.018  -4.8255 1.401e-06 ***
## immig2gen     5090.245   1213.987   4.1930 2.758e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


model2 <- lm(WSAL_VAL ~ Age + I(Age^2) + I(Age^3)+I(Age^4)
             + female + AfAm + Asian + Amindian + race_oth
             + Hispanic + educ_hs + educ_smcoll + educ_as + educ_bach +
 educ_adv
             + married + divwidsep + union_m + veteran + immigrant + im
mig2gen)

model3 <- lm(WSAL_VAL ~ Age + I(Age^2)
             + female + I(female*Age) + I(female*(Age^2))
             + AfAm + Asian + Amindian + race_oth
             + Hispanic + educ_hs + educ_smcoll + educ_as + educ_bach +
 educ_adv
             + married + divwidsep + union_m + veteran + immigrant + im
mig2gen)
summary(model2)
```

```
##
## Call:
## lm(formula = WSAL_VAL ~ Age + I(Age^2) + I(Age^3) + I(Age^4) +
##     female + AfAm + Asian + Amindian + race_oth + Hispanic +
##     educ_hs + educ_smcoll + educ_as + educ_bach + educ_adv +
##     married + divwidsep + union_m + veteran + immigrant + immig2gen)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -120964  -22659   -6391   11474 1072917
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.407e+05  1.506e+05   2.262   0.0237 *
## Age         -3.656e+04  1.587e+04  -2.304   0.0213 *
## I(Age^2)     1.489e+03  6.145e+02   2.422   0.0154 *
## I(Age^3)    -2.527e+01  1.037e+01  -2.435   0.0149 *
## I(Age^4)     1.545e-01  6.449e-02   2.396   0.0166 *
## female      -1.929e+04  5.648e+02 -34.157  < 2e-16 ***
## AfAm        -8.890e+03  9.255e+02  -9.606  < 2e-16 ***
## Asian       -1.531e+03  1.263e+03  -1.212   0.2255
## Amindian     2.003e+02  2.703e+03   0.074   0.9409
## race_oth    -4.044e+03  1.854e+03  -2.181   0.0292 *
## Hispanic    -6.789e+03  9.330e+02  -7.276 3.49e-13 ***
## educ_hs      9.140e+03  1.241e+03   7.365 1.80e-13 ***
## educ_smcoll  1.796e+04  1.326e+03  13.541  < 2e-16 ***
## educ_as      1.899e+04  1.400e+03  13.561  < 2e-16 ***
## educ_bach    3.924e+04  1.269e+03  30.922  < 2e-16 ***
## educ_adv     6.732e+04  1.356e+03  49.660  < 2e-16 ***
## married      5.827e+03  7.538e+02   7.730 1.09e-14 ***
## divwidsep    2.946e+02  1.008e+03   0.292   0.7700
## union_m     -2.547e+03  1.856e+03  -1.372   0.1701
## veteran      1.531e+02  1.180e+03   0.130   0.8968
## immigrant   -6.407e+03  1.210e+03  -5.297 1.19e-07 ***
## immig2gen    5.184e+03  1.133e+03   4.577 4.73e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 59100 on 47528 degrees of freedom
## Multiple R-squared:  0.1484, Adjusted R-squared:  0.1481
## F-statistic: 394.5 on 21 and 47528 DF,  p-value: < 2.2e-16
```

```
summary(model3)
```

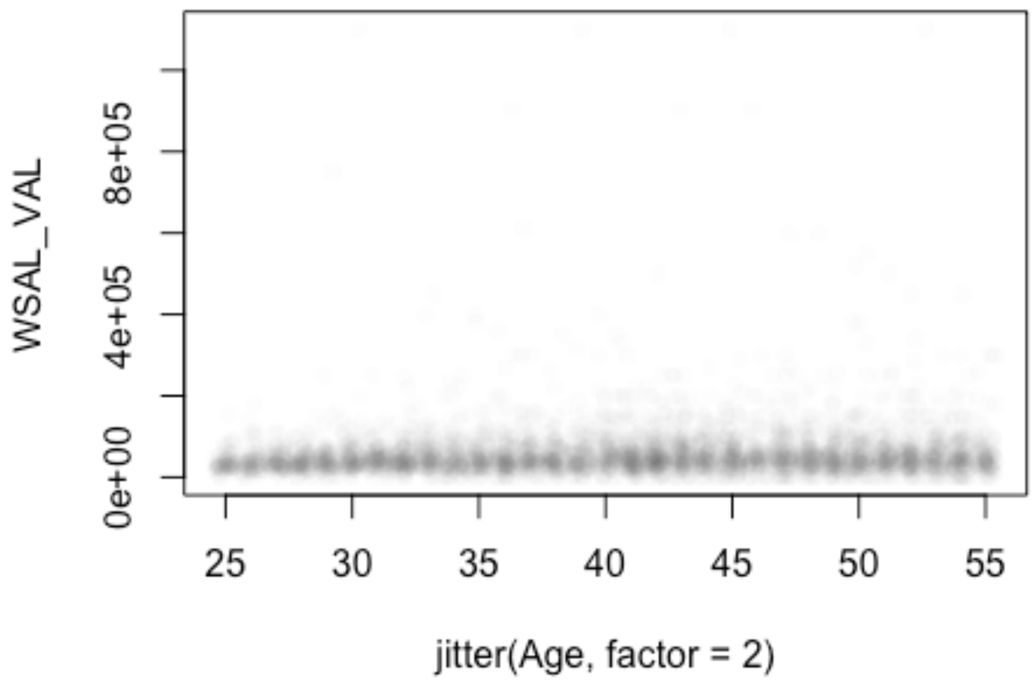
```
##
## Call:
## lm(formula = WSAL_VAL ~ Age + I(Age^2) + female + I(female *
##     Age) + I(female * (Age^2)) + AfAm + Asian + Amindian + race_oth
+
##     Hispanic + educ_hs + educ_smcoll + educ_as + educ_bach +
```

```
##     educ_adv + married + divwidsep + union_m + veteran + immigrant +

##     immig2gen)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -122605  -22595   -6176   11277 1073043
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            -57216.071   8515.010  -6.719 1.84e-11 ***
## Age                      3857.915    433.384   8.902  < 2e-16 ***
## I(Age^2)                  -36.507      5.319  -6.863 6.82e-12 ***
## female                  23707.136  12707.361   1.866  0.06210 .
## I(female * Age)         -1763.889    648.181  -2.721  0.00651 **
## I(female * (Age^2))        16.632      7.989   2.082  0.03737 *
## AfAm                    -8847.527    925.017  -9.565  < 2e-16 ***
## Asian                   -1465.503   1262.883  -1.160  0.24587
## Amindian                   94.265   2701.884   0.035  0.97217
## race_oth                -4023.447   1853.157  -2.171  0.02993 *
## Hispanic                -6767.933    932.594  -7.257 4.01e-13 ***
## educ_hs                  9184.964   1240.401   7.405 1.33e-13 ***
## educ_smcoll             17958.917   1325.863  13.545  < 2e-16 ***
## educ_as                 18968.396   1399.339  13.555  < 2e-16 ***
## educ_bach               39109.839   1268.435  30.833  < 2e-16 ***
## educ_adv                67073.889   1355.513  49.482  < 2e-16 ***
## married                  5685.914    751.387   7.567 3.88e-14 ***
## divwidsep                 405.647   1006.420   0.403  0.68691
## union_m                 -2510.200   1855.419  -1.353  0.17609
## veteran                  -285.831   1181.521  -0.242  0.80884
## immigrant               -6311.413   1209.235  -5.219 1.80e-07 ***
## immig2gen                5138.880   1132.110   4.539 5.66e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 59070 on 47528 degrees of freedom
## Multiple R-squared:  0.1492, Adjusted R-squared:  0.1488
## F-statistic: 396.8 on 21 and 47528 DF,  p-value: < 2.2e-16
```

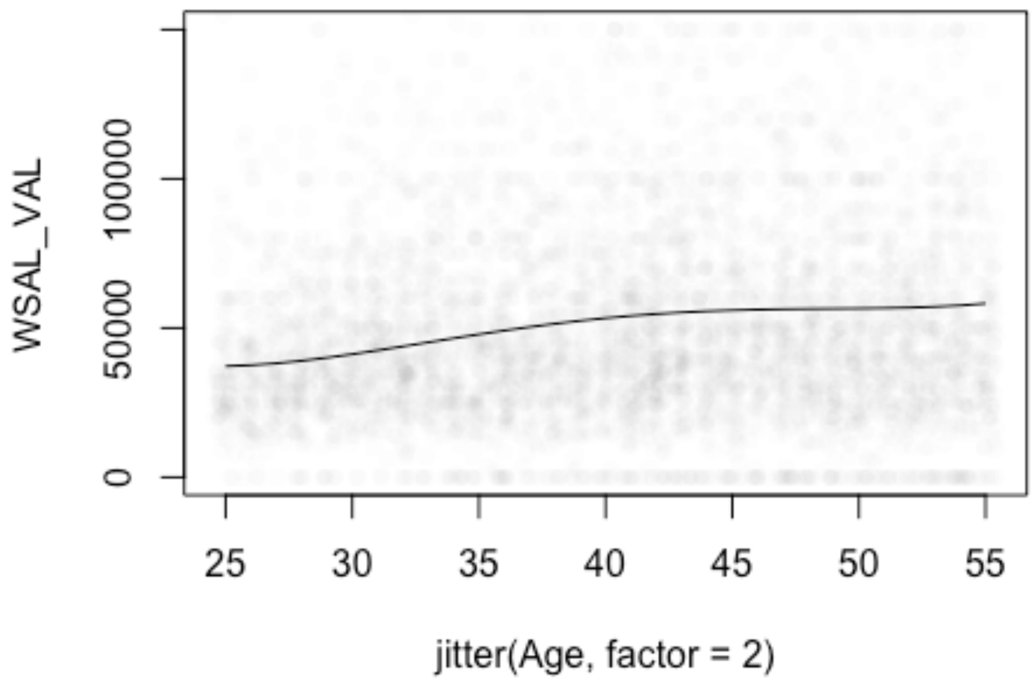
```r
anova(model1,model2,model3)
```

```
## Analysis of Variance Table
##
## Model 1: WSAL_VAL ~ Age + female + AfAm + Asian + Amindian + race_ot
h +
##     Hispanic + educ_hs + educ_smcoll + educ_as + educ_bach +
##     educ_adv + married + divwidsep + union_m + veteran + immigrant +

##     immig2gen
```

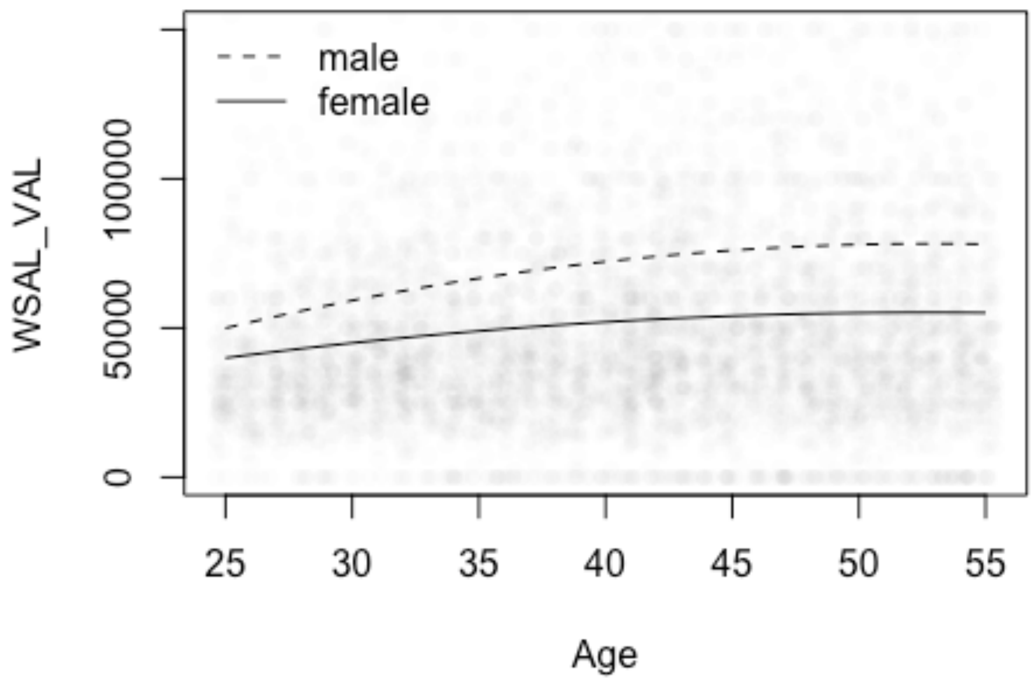
```
## Model 2: WSAL_VAL ~ Age + I(Age^2) + I(Age^3) + I(Age^4) + female +
AfAm +
##     Asian + Amindian + race_oth + Hispanic + educ_hs + educ_smcoll +

##     educ_as + educ_bach + educ_adv + married + divwidsep + union_m +

##     veteran + immigrant + immig2gen
## Model 3: WSAL_VAL ~ Age + I(Age^2) + female + I(female * Age) + I(fe
male *
##     (Age^2)) + AfAm + Asian + Amindian + race_oth + Hispanic +
##     educ_hs + educ_smcoll + educ_as + educ_bach + educ_adv +
##     married + divwidsep + union_m + veteran + immigrant + immig2gen
##   Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
## 1  47531 1.6619e+14
## 2  47528 1.6598e+14  3 2.0543e+11 19.608 1.067e-12 ***
## 3  47528 1.6584e+14  0 1.4402e+11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

NNobs <- length(WSAL_VAL)
set.seed(12345)

graph_obs <- (runif(NNobs) < 0.1)
dat_graph <-subset(dat_use,graph_obs)

plot(WSAL_VAL ~ jitter(Age, factor = 2), pch = 16, col = rgb(0.5, 0.5,
0.5, alpha = 0.02), data = dat_graph)
```




```r
# ^^ that looks like crap since Wages are sooooooooo skew!  So try to fi
nd ylim = c(0, ??)
plot(WSAL_VAL ~ jitter(Age, factor = 2), pch = 16, col = rgb(0.5, 0.5,
0.5, alpha = 0.02), ylim = c(0,150000), data = dat_graph)

# to plot the predicted values might want to do something like, lines(f
itted.values(model2) ~ Age)
# but that will plot ALLLLL the values, which is 4500 too many and look
s awful
# so back to this,
to_be_predicted2 <- data.frame(Age = 25:55, female = 1, AfAm = 0, Asian
 = 0, Amindian = 1, race_oth = 1,
                               Hispanic = 1, educ_hs = 0, educ_smcoll =
 0, educ_as = 0, educ_bach = 1, educ_adv = 0,
                               married = 0, divwidsep =0, union_m = 0,
veteran = 0, immigrant = 0, immig2gen = 1)
to_be_predicted2$yhat <- predict(model2, newdata = to_be_predicted2)

lines(yhat ~ Age, data = to_be_predicted2)
```




```
# now compare model3
to_be_predicted3m <- data.frame(Age = 25:55, female = 0, AfAm = 0, Asia
n = 0, Amindian = 1, race_oth = 1,
                                Hispanic = 1, educ_hs = 0, educ_smcoll
= 0, educ_as = 0, educ_bach = 1, educ_adv = 0,
                                married = 0, divwidsep =0, union_m = 0,
 veteran = 0, immigrant = 0, immig2gen = 1)
to_be_predicted3m$yhat <- predict(model3, newdata = to_be_predicted3m)

to_be_predicted3f <- data.frame(Age = 25:55, female = 1, AfAm = 0, Asia
n = 0, Amindian = 1, race_oth = 1,
                                Hispanic = 1, educ_hs = 0, educ_smcoll
= 0, educ_as = 0, educ_bach = 1, educ_adv = 0,
                                married = 0, divwidsep =0, union_m = 0,
 veteran = 0, immigrant = 0, immig2gen = 1)
to_be_predicted3f$yhat <- predict(model3, newdata = to_be_predicted3f)

plot(WSAL_VAL ~ jitter(Age, factor = 2), pch = 16, col = rgb(0.5, 0.5,
0.5, alpha = 0.02), ylim = c(0,150000), xlab = "Age", data = dat_graph)
lines(yhat ~ Age, data = to_be_predicted3f)
lines(yhat ~ Age, data = to_be_predicted3m, lty = 2)
legend("topleft", c("male", "female"), lty = c(2,1), bty = "n")
```




```r
detach(dat_use)
```

(4) **Next add some multilevel effects on some of those factors like industry, occupation, and/or state. Again show nice output.**

For the above regressions, industry factors like: industry, occupation, and state, were not taken into account and left out of the regression. In other words the sampled individuals all had the same value when it came to these multilevel effects. In multilevel model, these factors like industry are assigned a varying intercept term, therein not constraining the intercept and allowing from variation high to low. So with this model those we get an average of those that started with the same value or as called in class "pooled" and those individuals with differently valued factors referred to in class as "no-pooled". Why use this over multiple regression? While also a good model it treats the variables as independent an ignoring groupings may cause one underestimation of in standard error of regression coefficients which will give you a higher statistical significance. For this model we decided to add multilevel effects of industry. To accomplish download the lme4 package for R.and use code lmer (instead lm that we have been previously using) as demonstrated below we start with nonlinear regression and build on that:

Raw Data:

```
> rm(list = ls(all = TRUE))
> load("~/cps_mar2013.RData")
> attach(dat_CPSMar2013)
> use_varb <- (Age >= 25) & (Age <= 55) & work_fullt & work_50wks
> dat_use <- subset(dat_CPSMar2013,use_varb)
> detach(dat_CPSMar2013)
> educ_indx <- as.factor(educ_nohs + 2*educ_hs + 3*educ_smcoll + 4*educ
_as + 5*educ_bach + 6*educ_adv)
> attach(dat_use)

> by(WSAL_VAL, A_DTOCC, summary)
A_DTOCC: 1
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0   40000   65000   81130  100000 1100000
----------------------------------------------------------------
A_DTOCC: 2
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0   40000   57000   73200   85000 1100000
----------------------------------------------------------------
A_DTOCC: 3
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0   55000   75000   82390  100000 1100000
----------------------------------------------------------------
A_DTOCC: 4
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0   55000   75000   88040  100000 1100000
----------------------------------------------------------------
A_DTOCC: 5
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0   48000   65000   76220   94000 1100000
----------------------------------------------------------------
A_DTOCC: 6
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0   31500   42000   44940   55000  200000
----------------------------------------------------------------
```

```
A_DTOCC: 7
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0   43700   75000  108200  130000 1100000 
----------------------------------------------------------------- 
A_DTOCC: 8
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0   34000   46000   50310   60000  600000 
----------------------------------------------------------------- 
A_DTOCC: 9
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0   28950   45000   55530   72000  750000 
----------------------------------------------------------------- 
A_DTOCC: 10
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0   42000   60000   81040   84000 1100000 
----------------------------------------------------------------- 
A_DTOCC: 11
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0   21000   27000   29600   35000  185000 
----------------------------------------------------------------- 
A_DTOCC: 12
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0   36000   55000   61580   80000 1100000 
----------------------------------------------------------------- 
A_DTOCC: 13
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0   17500   24000   27500   32000  192000 
----------------------------------------------------------------- 
A_DTOCC: 14
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0   16000   24000   27360   35000 1100000 
----------------------------------------------------------------- 
A_DTOCC: 15
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0    5000   20400   24640   33700  900000 
----------------------------------------------------------------- 
A_DTOCC: 16
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0   25000   40000   57380   68000 1100000 
----------------------------------------------------------------- 
A_DTOCC: 17
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0   27000   35790   41680   50000 1100000 
----------------------------------------------------------------- 
A_DTOCC: 18
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0   18000   25000   28670   34000  149100 
----------------------------------------------------------------- 
A_DTOCC: 19
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0   24000   38000   44370   60000 1100000 
----------------------------------------------------------------- 
A_DTOCC: 20
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0   30000   45000   49810   62000 1100000 
----------------------------------------------------------------- 
A_DTOCC: 21
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0   25000   35100   41760   50000 1100000 
---------------------------------------------------------------- 
A_DTOCC: 22
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0   25000   38000   42860   52000 1100000 
> plot(as.factor(female) ~ A_DTOCC)
> model1 <- lm(WSAL_VAL ~ Age + female + AfAm + Asian + Amindian + race
_oth
+ Hispanic + educ_hs + educ_smcoll + educ_as + educ_bach + educ_adv
+ married + divwidsep + union_m + veteran + immigrant + immig2gen, data
 = dat_use)
> summary(model1)

Call:
lm(formula = WSAL_VAL ~ Age + female + AfAm + Asian + Amindian + 
    race_oth + Hispanic + educ_hs + educ_smcoll + educ_as + educ_bach + 

    educ_adv + married + divwidsep + union_m + veteran + immigrant + 
    immig2gen, data = dat_use)

Residuals:
    Min      1Q  Median      3Q     Max 
-123822  -22631   -6241   11283 1075277 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   6656.73    1801.46   3.695  0.00022 ***
Age            711.93      33.73  21.108  < 2e-16 ***
female      -19334.07     565.14 -34.211  < 2e-16 ***
AfAm         -8679.40     925.52  -9.378  < 2e-16 ***
Asian        -1625.51    1264.07  -1.286  0.19847    
Amindian       329.19    2704.50   0.122  0.90312    
race_oth     -4032.07    1855.04  -2.174  0.02974 *  
Hispanic     -6719.13     933.50  -7.198 6.21e-13 ***
educ_hs       9117.64    1241.65   7.343 2.12e-13 ***
educ_smcoll  17993.55    1327.20  13.557  < 2e-16 ***
educ_as      19046.51    1400.74  13.597  < 2e-16 ***
educ_bach    39285.83    1269.56  30.944  < 2e-16 ***
educ_adv     67575.27    1355.90  49.838  < 2e-16 ***
married       6767.08     738.77   9.160  < 2e-16 ***
divwidsep     1228.63     997.57   1.232  0.21809    
union_m      -2575.03    1857.28  -1.386  0.16562    
veteran        182.37    1180.67   0.154  0.87724    
immigrant    -6094.72    1209.60  -5.039 4.71e-07 ***
immig2gen     5090.24    1133.14   4.492 7.07e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 59130 on 47531 degrees of freedom
Multiple R-squared:  0.1474,	Adjusted R-squared:  0.1471 
F-statistic: 456.5 on 18 and 47531 DF,  p-value: < 2.2e-16

> coeftest(model1,vcovHC)

t test of coefficients:
```

```
                Estimate Std. Error  t value  Pr(>|t|)    
(Intercept)   6656.733   1433.365   4.6441 3.424e-06 ***
Age            711.935     31.137  22.8649 < 2.2e-16 ***
female      -19334.071    539.294 -35.8507 < 2.2e-16 ***
AfAm         -8679.399    646.035 -13.4349 < 2.2e-16 ***
Asian        -1625.510   1310.443  -1.2404  0.214823    
Amindian       329.194   3471.314   0.0948  0.924448    
race_oth     -4032.066   1490.067  -2.7060  0.006813 ** 
Hispanic     -6719.129    933.994  -7.1940 6.386e-13 ***
educ_hs       9117.636    787.394  11.5795 < 2.2e-16 ***
educ_smcoll  17993.547    915.721  19.6496 < 2.2e-16 ***
educ_as      19046.514    915.212  20.8110 < 2.2e-16 ***
educ_bach    39285.835    945.628  41.5447 < 2.2e-16 ***
educ_adv     67575.269   1468.396  46.0198 < 2.2e-16 ***
married       6767.082    603.297  11.2168 < 2.2e-16 ***
divwidsep     1228.629    838.934   1.4645  0.143061    
union_m      -2575.027   1541.328  -1.6707  0.094796 .  
veteran        182.371   1313.661   0.1388  0.889588    
immigrant    -6094.717   1263.018  -4.8255 1.401e-06 ***
immig2gen     5090.245   1213.987   4.1930 2.758e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> model_continuousAge <- lm(WSAL_VAL ~ Age, data = dat_use)
> summary(model_continuousAge)

Call:
lm(formula = WSAL_VAL ~ Age, data = dat_use)

Residuals:
    Min      1Q  Median      3Q     Max 
 -68318  -28318  -12424   11802 1051027 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  25755.2     1416.7   18.18   <2e-16 ***
Age            773.9       34.0   22.76   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 63680 on 47548 degrees of freedom
Multiple R-squared:  0.01078,	Adjusted R-squared:  0.01076 
F-statistic: 518.2 on 1 and 47548 DF,  p-value: < 2.2e-16

> Age_factr <- cut(dat_use$Age,breaks=25:55)
> model_factrAge <-lm(WSAL_VAL ~ Age_factr, data = dat_use)
> summary(model_factrAge)

Call:
lm(formula = WSAL_VAL ~ Age_factr, data = dat_use)

Residuals:
    Min      1Q  Median      3Q     Max 
 -65179  -28920  -12132   11828 1052178 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
```

```
(Intercept)          39200       1972  19.880  < 2e-16 ***
Age_factr(26,27]      2051       2707   0.758 0.448570    
Age_factr(27,28]      4737       2663   1.779 0.075265 .  
Age_factr(28,29]      5586       2653   2.106 0.035250 *  
Age_factr(29,30]      8620       2594   3.323 0.000890 ***
Age_factr(30,31]     11188       2597   4.308 1.65e-05 ***
Age_factr(31,32]      9781       2561   3.819 0.000134 ***
Age_factr(32,33]     12329       2570   4.798 1.61e-06 ***
Age_factr(33,34]     14386       2574   5.590 2.28e-08 ***
Age_factr(34,35]     17295       2564   6.744 1.56e-11 ***
Age_factr(35,36]     17849       2557   6.980 2.98e-12 ***
Age_factr(36,37]     19758       2553   7.738 1.03e-14 ***
Age_factr(37,38]     20698       2576   8.035 9.56e-16 ***
Age_factr(38,39]     19960       2585   7.722 1.16e-14 ***
Age_factr(39,40]     19830       2532   7.831 4.94e-15 ***
Age_factr(40,41]     20756       2507   8.278  < 2e-16 ***
Age_factr(41,42]     22720       2476   9.177  < 2e-16 ***
Age_factr(42,43]     25495       2508  10.163  < 2e-16 ***
Age_factr(43,44]     23635       2518   9.388  < 2e-16 ***
Age_factr(44,45]     25979       2507  10.364  < 2e-16 ***
Age_factr(45,46]     21512       2522   8.531  < 2e-16 ***
Age_factr(46,47]     22932       2515   9.117  < 2e-16 ***
Age_factr(47,48]     21104       2536   8.323  < 2e-16 ***
Age_factr(48,49]     24907       2498   9.970  < 2e-16 ***
Age_factr(49,50]     24303       2496   9.735  < 2e-16 ***
Age_factr(50,51]     21282       2532   8.404  < 2e-16 ***
Age_factr(51,52]     25227       2547   9.905  < 2e-16 ***
Age_factr(52,53]     24096       2539   9.492  < 2e-16 ***
Age_factr(53,54]     23972       2585   9.275  < 2e-16 ***
Age_factr(54,55]     24065       2578   9.333  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 64200 on 46509 degrees of freedom
  (1011 observations deleted due to missingness)
Multiple R-squared:  0.0117,   Adjusted R-squared:  0.01108 
F-statistic: 18.98 on 29 and 46509 DF,  p-value: < 2.2e-16

> plot(coef(model_factrAge))
> model2 <- lm(WSAL_VAL ~ Age + female + AfAm + Asian + Amindian + race
_oth
+               + Hispanic + educ_hs + educ_smcoll + educ_as + educ_bach
 + educ_adv
+               + married + divwidsep + union_m + veteran + immigrant +
immig2gen
+               + as.factor(A_DTOCC), data = dat_use)
> summary(model2)

Call:
lm(formula = WSAL_VAL ~ Age + female + AfAm + Asian + Amindian + 
    race_oth + Hispanic + educ_hs + educ_smcoll + educ_as + educ_bach + 

    educ_adv + married + divwidsep + union_m + veteran + immigrant + 
    immig2gen + as.factor(A_DTOCC), data = dat_use)

Residuals:
    Min      1Q  Median      3Q     Max 
```

```
-140733  -21102   -5063   11130 1068669

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)           30203.97    1995.41  15.137  < 2e-16 ***
Age                     666.62      33.19  20.086  < 2e-16 ***
female               -17721.52     624.17 -28.392  < 2e-16 ***
AfAm                  -7054.01     913.19  -7.725 1.14e-14 ***
Asian                 -2943.64    1244.26  -2.366 0.017997 *  
Amindian               1977.22    2650.75   0.746 0.455725    
race_oth              -3775.58    1818.08  -2.077 0.037836 *  
Hispanic              -4755.59     917.38  -5.184 2.18e-07 ***
educ_hs                5873.18    1231.55   4.769 1.86e-06 ***
educ_smcoll           11729.11    1336.82   8.774  < 2e-16 ***
educ_as               10869.74    1418.67   7.662 1.87e-14 ***
educ_bach             29106.51    1336.41  21.780  < 2e-16 ***
educ_adv              57147.55    1472.32  38.815  < 2e-16 ***
married                5713.11     726.91   7.859 3.94e-15 ***
divwidsep               695.49     978.85   0.711 0.477389    
union_m                4127.95    1837.76   2.246 0.024697 *  
veteran                -270.72    1162.12  -0.233 0.815799    
immigrant             -4291.58    1192.16  -3.600 0.000319 ***
immig2gen              4037.26    1110.90   3.634 0.000279 ***
as.factor(A_DTOCC)2   -5301.02    1322.13  -4.009 6.10e-05 ***
as.factor(A_DTOCC)3   -3175.84    1518.65  -2.091 0.036513 *  
as.factor(A_DTOCC)4    -176.31    1829.15  -0.096 0.923211    
as.factor(A_DTOCC)5  -18261.81    2583.23  -7.069 1.58e-12 ***
as.factor(A_DTOCC)6  -37418.00    2050.68 -18.247  < 2e-16 ***
as.factor(A_DTOCC)7   11670.85    2286.04   5.105 3.32e-07 ***
as.factor(A_DTOCC)8  -37481.46    1392.01 -26.926  < 2e-16 ***
as.factor(A_DTOCC)9  -22897.47    2186.53 -10.472  < 2e-16 ***
as.factor(A_DTOCC)10   2634.96    1335.58   1.973 0.048514 *  
as.factor(A_DTOCC)11 -23432.10    2108.47 -11.113  < 2e-16 ***
as.factor(A_DTOCC)12 -11738.99    1767.91  -6.640 3.17e-11 ***
as.factor(A_DTOCC)13 -29852.83    1743.67 -17.121  < 2e-16 ***
as.factor(A_DTOCC)14 -32581.18    1754.78 -18.567  < 2e-16 ***
as.factor(A_DTOCC)15 -32947.19    1852.60 -17.784  < 2e-16 ***
as.factor(A_DTOCC)16 -13052.48    1145.28 -11.397  < 2e-16 ***
as.factor(A_DTOCC)17 -19524.62    1092.63 -17.869  < 2e-16 ***
as.factor(A_DTOCC)18 -31852.89    3553.06  -8.965  < 2e-16 ***
as.factor(A_DTOCC)19 -22729.55    1469.89 -15.463  < 2e-16 ***
as.factor(A_DTOCC)20 -20537.64    1516.74 -13.541  < 2e-16 ***
as.factor(A_DTOCC)21 -22071.98    1338.83 -16.486  < 2e-16 ***
as.factor(A_DTOCC)22 -22550.08    1394.51 -16.171  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 57940 on 47510 degrees of freedom
Multiple R-squared:  0.1818,	Adjusted R-squared:  0.1811 
F-statistic: 270.6 on 39 and 47510 DF,  p-value: < 2.2e-16

> coeftest(model2,vcovHC)

t test of coefficients:

                       Estimate Std. Error  t value  Pr(>|t|)    
(Intercept)           30203.975   1632.890  18.4972 < 2.2e-16 ***
```

```
Age                         666.620     30.352  21.9631 < 2.2e-16 ***
female                   -17721.521    656.971 -26.9746 < 2.2e-16 ***
AfAm                      -7054.011    629.251 -11.2102 < 2.2e-16 ***
Asian                     -2943.636   1287.068  -2.2871 0.0221951 *
Amindian                   1977.222   3473.532   0.5692 0.5692059
race_oth                  -3775.584   1461.996  -2.5825 0.0098121 **
Hispanic                  -4755.586    918.394  -5.1782 2.250e-07 ***
educ_hs                    5873.180    784.225   7.4892 7.052e-14 ***
educ_smcoll               11729.113    937.277  12.5140 < 2.2e-16 ***
educ_as                   10869.738    983.102  11.0566 < 2.2e-16 ***
educ_bach                 29106.509    986.951  29.4913 < 2.2e-16 ***
educ_adv                  57147.552   1500.473  38.0863 < 2.2e-16 ***
married                    5713.111    596.326   9.5805 < 2.2e-16 ***
divwidsep                   695.487    823.932   0.8441 0.3986136
union_m                    4127.949   1447.439   2.8519 0.0043478 **
veteran                    -270.719   1300.239  -0.2082 0.8350679
immigrant                 -4291.581   1243.908  -3.4501 0.0005609 ***
immig2gen                  4037.257   1190.203   3.3921 0.0006942 ***
as.factor(A_DTOCC)2       -5301.023   1616.573  -3.2792 0.0010419 **
as.factor(A_DTOCC)3       -3175.843   1573.216  -2.0187 0.0435245 *
as.factor(A_DTOCC)4        -176.312   2352.350  -0.0750 0.9402536
as.factor(A_DTOCC)5      -18261.809   2626.525  -6.9528 3.626e-12 ***
as.factor(A_DTOCC)6      -37418.005   1325.976 -28.2192 < 2.2e-16 ***
as.factor(A_DTOCC)7       11670.853   4441.142   2.6279 0.0085943 **
as.factor(A_DTOCC)8      -37481.459   1182.849 -31.6874 < 2.2e-16 ***
as.factor(A_DTOCC)9      -22897.473   2111.697 -10.8432 < 2.2e-16 ***
as.factor(A_DTOCC)10       2634.957   1946.410   1.3538 0.1758218
as.factor(A_DTOCC)11 -23432.101   1089.801 -21.5013 < 2.2e-16 ***
as.factor(A_DTOCC)12 -11738.985   1419.113  -8.2721 < 2.2e-16 ***
as.factor(A_DTOCC)13 -29852.826   1064.587 -28.0417 < 2.2e-16 ***
as.factor(A_DTOCC)14 -32581.177   1327.216 -24.5485 < 2.2e-16 ***
as.factor(A_DTOCC)15 -32947.190   1432.454 -23.0005 < 2.2e-16 ***
as.factor(A_DTOCC)16 -13052.483   1485.787  -8.7849 < 2.2e-16 ***
as.factor(A_DTOCC)17 -19524.623   1036.320 -18.8403 < 2.2e-16 ***
as.factor(A_DTOCC)18 -31852.888   1478.723 -21.5408 < 2.2e-16 ***
as.factor(A_DTOCC)19 -22729.552   1369.944 -16.5916 < 2.2e-16 ***
as.factor(A_DTOCC)20 -20537.641   1274.380 -16.1158 < 2.2e-16 ***
as.factor(A_DTOCC)21 -22071.983   1283.961 -17.1905 < 2.2e-16 ***
as.factor(A_DTOCC)22 -22550.076   1216.234 -18.5409 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> modelmm2 <- lmer(WSAL_VAL ~ Age + female + AfAm + Asian + Amindian +
race_oth
+                   + Hispanic + educ_hs + educ_smcoll + educ_as + educ_
bach + educ_adv
+                   + married + divwidsep + union_m + veteran + immigran
t + immig2gen
+                   + as.factor(A_DTOCC) + (1 | as.factor(A_DTOCC)), dat
_use)
> summary(modelmm2)
Linear mixed model fit by REML ['lmerMod']
Formula: WSAL_VAL ~ Age + female + AfAm + Asian + Amindian + race_oth +

    Hispanic + educ_hs + educ_smcoll + educ_as + educ_bach +
    educ_adv + married + divwidsep + union_m + veteran + immigrant +
    immig2gen + as.factor(A_DTOCC) + (1 | as.factor(A_DTOCC))
```

```
   Data: dat_use

REML criterion at convergence: 1177242

Scaled residuals: 
    Min      1Q  Median      3Q     Max 
-2.4290 -0.3642 -0.0874  0.1921 18.4447 

Random effects:
 Groups             Name        Variance  Std.Dev.
 as.factor(A_DTOCC) (Intercept) 2.289e+08 15128   
 Residual                       3.357e+09 57939   
Number of obs: 47550, groups:  as.factor(A_DTOCC), 22

Fixed effects:
                      Estimate Std. Error t value
(Intercept)           30203.97   15259.51    1.98
Age                     666.62      33.19   20.09
female               -17721.52     624.17  -28.39
AfAm                  -7054.01     913.19   -7.72
Asian                 -2943.64    1244.26   -2.37
Amindian               1977.22    2650.75    0.75
race_oth              -3775.58    1818.08   -2.08
Hispanic              -4755.59     917.38   -5.18
educ_hs                5873.18    1231.55    4.77
educ_smcoll           11729.11    1336.82    8.77
educ_as               10869.74    1418.67    7.66
educ_bach             29106.51    1336.41   21.78
educ_adv              57147.55    1472.32   38.81
married                5713.11     726.91    7.86
divwidsep               695.49     978.85    0.71
union_m                4127.95    1837.76    2.25
veteran                -270.72    1162.12   -0.23
immigrant             -4291.58    1192.16   -3.60
immig2gen              4037.26    1110.90    3.63
as.factor(A_DTOCC)2   -5301.02   21435.71   -0.25
as.factor(A_DTOCC)3   -3175.84   21448.73   -0.15
as.factor(A_DTOCC)4    -176.31   21472.95   -0.01
as.factor(A_DTOCC)5  -18261.81   21550.29   -0.85
as.factor(A_DTOCC)6  -37418.00   21492.95   -1.74
as.factor(A_DTOCC)7   11670.85   21516.68    0.54
as.factor(A_DTOCC)8  -37481.46   21440.14   -1.75
as.factor(A_DTOCC)9  -22897.47   21506.34   -1.06
as.factor(A_DTOCC)10   2634.96   21436.55    0.12
as.factor(A_DTOCC)11 -23432.10   21498.54   -1.09
as.factor(A_DTOCC)12 -11738.99   21467.82   -0.55
as.factor(A_DTOCC)13 -29852.83   21465.84   -1.39
as.factor(A_DTOCC)14 -32581.18   21466.74   -1.52
as.factor(A_DTOCC)15 -32947.19   21474.96   -1.53
as.factor(A_DTOCC)16 -13052.48   21425.53   -0.61
as.factor(A_DTOCC)17 -19524.62   21422.78   -0.91
as.factor(A_DTOCC)18 -31852.89   21687.92   -1.47
as.factor(A_DTOCC)19 -22729.55   21445.33   -1.06
as.factor(A_DTOCC)20 -20537.64   21448.60   -0.96
as.factor(A_DTOCC)21 -22071.98   21436.75   -1.03
as.factor(A_DTOCC)22 -22550.08   21440.30   -1.05
```

```
Correlation matrix not shown by default, as p = 40 > 12.
Use print(x, correlation=TRUE)  or
        vcov(x)        if you need it
```

So here we see the random intercept added into modelmm2 of salary and the detaile d occupation recode on random and fixed effects. Here the model may be fitted by m inimizing the marginal density for Y with respect to the fixed effects.

Random effects:

Groups Name	Variance	Std.Dev.
as.factor(A_DTOCC) (Intercept)	2.289e+08	15128
Residual	3.357e+09	57939

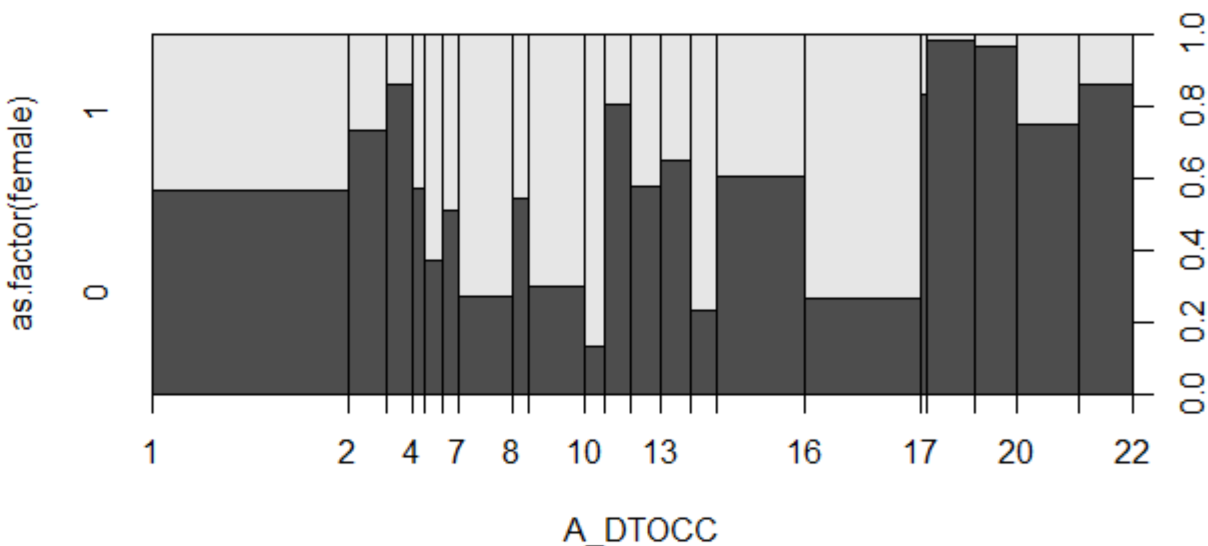
Number of obs: 47550, groups: as.factor(A_DTOCC), 22

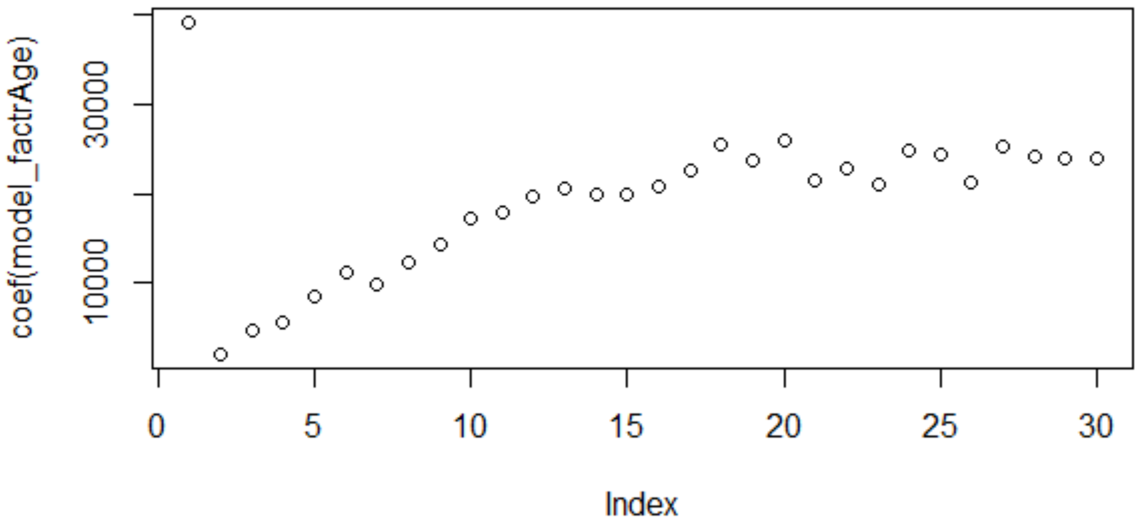
Below is the t value for the fixed effects for each individual intercept:
Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	30203.97	15259.51	1.98
Age	666.62	33.19	20.09
female	-17721.52	624.17	-28.39
AfAm	-7054.01	913.19	-7.72
Asian	-2943.64	1244.26	-2.37
Amindian	1977.22	2650.75	0.75
race_oth	-3775.58	1818.08	-2.08
Hispanic	-4755.59	917.38	-5.18
educ_hs	5873.18	1231.55	4.77
educ_smcoll	11729.11	1336.82	8.77
educ_as	10869.74	1418.67	7.66
educ_bach	29106.51	1336.41	21.78
educ_adv	57147.55	1472.32	38.81
married	5713.11	726.91	7.86
divwidsep	695.49	978.85	0.71
union_m	4127.95	1837.76	2.25
veteran	-270.72	1162.12	-0.23
immigrant	-4291.58	1192.16	-3.60
immig2gen	4037.26	1110.90	3.63
as.factor(A_DTOCC)2	-5301.02	21435.71	-0.25
as.factor(A_DTOCC)3	-3175.84	21448.73	-0.15
as.factor(A_DTOCC)4	-176.31	21472.95	-0.01
as.factor(A_DTOCC)5	-18261.81	21550.29	-0.85
as.factor(A_DTOCC)6	-37418.00	21492.95	-1.74
as.factor(A_DTOCC)7	11670.85	21516.68	0.54
as.factor(A_DTOCC)8	-37481.46	21440.14	-1.75
as.factor(A_DTOCC)9	-22897.47	21506.34	-1.06
as.factor(A_DTOCC)10	2634.96	21436.55	0.12
as.factor(A_DTOCC)11	-23432.10	21498.54	-1.09

as.factor(A_DTOCC)12	-11738.99	21467.82	-0.55
as.factor(A_DTOCC)13	-29852.83	21465.84	-1.39
as.factor(A_DTOCC)14	-32581.18	21466.74	-1.52
as.factor(A_DTOCC)15	-32947.19	21474.96	-1.53
as.factor(A_DTOCC)16	-13052.48	21425.53	-0.61
as.factor(A_DTOCC)17	-19524.62	21422.78	-0.91
as.factor(A_DTOCC)18	-31852.89	21687.92	-1.47
as.factor(A_DTOCC)19	-22729.55	21445.33	-1.06
as.factor(A_DTOCC)20	-20537.64	21448.60	-0.96
as.factor(A_DTOCC)21	-22071.98	21436.75	-1.03
as.factor(A_DTOCC)22	-22550.08	21440.30	-1.05

Charts:




coef(model_factrAge)
10000
30000
0
5
10
15
20
25
30
Index

Group members: Keyi Long, Crystal Hernandaz, Emmanuel Monroy.

Question 2 - Nonlinear Regression Model of Log Wages and Salary

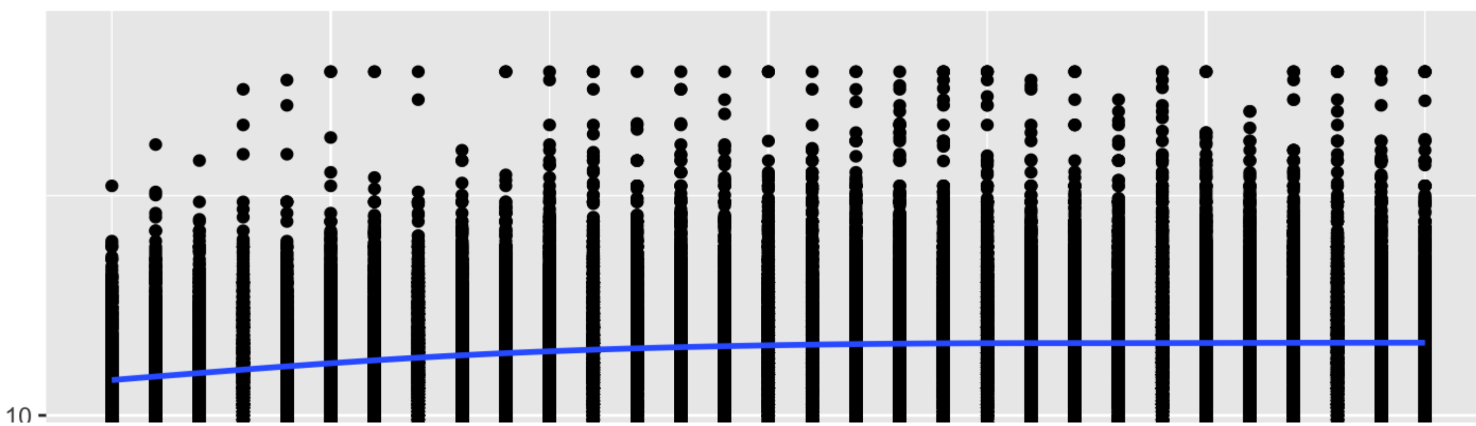
Depedent Varibles: Log Wages and Salary; Prime Age (25<=Age<=55)

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

Regressor	(1)	(2)	(3)	(4)
Age	0.0103909*** (0.0003413)	5.440e-02*** (3.334e-03)	1.034e-01*** (2.434e-02)	-1.088e-01 (1.600e-01)
Age^2		-5.421e-04*** (4.085e-05)	-1.794e-03** (6.171e-04)	6.484e-03 (6.199e-03)
Age^3			1.036e-05* (5.096e-06)	-1.300e-04 (1.047e-04)
Age^4				8.741e-07 (6.513e-07)
female	-0.3066790*** 0.0057144	-3.060e-01*** (5.704e-03)	-3.059e-01*** (5.704e-03)	-3.059e-01*** (5.704e-03)
Afam(African)	-0.1348315*** (0.0093219)	-1.386e-01*** (9.308e-03)	-1.389e-01*** (9.309e-03)	-1.390e-01*** (9.309e-03)
Asian	0.0091999 (0.0127948)	1.049e-02 (1.277e-02)	1.026e-02 (1.277e-02)	1.025e-02 (1.277e-02)

Amindian	-0.0690166* (0.0273319)	-7.226e-02** (2.728e-02)	-7.294e-02** (2.728e-02)	-7.276e-02** (2.728e-02)
race_oth	-0.0521587** (0.0187291)	-5.246e-02** (1.869e-02)	-5.271e-02** (1.869e-02)	-5.270e-02** (1.869e-02)
Hispanic	-0.1183893*** (0.0094346)	-1.195e-01*** (9.417e-03)	-1.194e-01*** (9.417e-03)	-1.194e-01*** (9.417e-03)
educ_hs	0.2552920*** (0.0126989)	2.559e-01*** (1.267e-02)	2.562e-01*** (1.268e-02)	2.562e-01*** (1.268e-02)
educ_smcoll	0.4169565*** (0.0135501)	4.164e-01*** (1.352e-02)	4.166e-01*** (1.352e-02)	4.166e-01*** (1.352e-02)
educ_as	0.4656438*** (0.0142824)	4.643e-01*** (1.426e-02)	4.644e-01*** (1.426e-02)	4.645e-01*** (1.425e-02)
educ_bach	0.7622496*** (0.0129570)	7.613e-01*** (1.293e-02)	7.616e-01*** (1.293e-02)	7.617e-01*** (1.293e-02)
educ_adv	1.0545368*** (0.0137985)	1.049e+00*** (1.378e-02)	1.049e+00*** (1.378e-02)	1.049e+00*** (1.378e-02)
married	0.1445854*** (0.0074456)	1.258e-01*** (7.566e-03)	1.246e-01*** (7.586e-03)	1.248e-01*** (7.587e-03)
dividsep	0.0519897*** (0.0100821)	3.345e-02*** (1.016e-02)	3.255e-02** (1.017e-02)	3.268e-02** (1.017e-02)
union_m	0.0307778 (0.0184083)	3.154e-02 (1.837e-02)	3.128e-02 (1.837e-02)	3.115e-02 (1.837e-02)
veteran	0.0297534* (0.0119379)	2.959e-02* (1.191e-02)	2.985e-02* (1.192e-02)	2.989e-02* (1.192e-02)
immigrant	-0.1456057*** (0.0122177)	-1.513e-01*** 1.220e-02	-1.513e-01*** (1.220e-02)	-1.513e-01*** (1.220e-02)
immig2gen	0.0747945*** (0.0114220)	7.656e-02*** (1.140e-02)	7.663e-02*** (1.140e-02)	7.665e-02*** (1.140e-02)
Intercept	9.8508146*** (0.0183002)	9.015e+00*** (6.560e-02)	8.396e+00*** (3.116e-01)	1.039e+01*** (1.517e+00)

F-Statistics and p-Values on Joint Hypotheses				
(a) Age, Age^2 = 0		553.17 (< 2.2e-16) ***		
(c) Age^2, Age^3 = 0			90.115 (< 2.2e-16) ***	7.9363 (0.000358)***
(d) Age^2, Age^3 , Age^4 = 0				60.678 < 2.2e-16 ***
F statistic for the whole model	1072 (< 2.2e-16)	1029 (< 2.2e-16)	977.5 (< 2.2e-16)	931 (< 2.2e-16)
SER	0.5858	0.5846	0.5846	0.5846
R^2	0.2977	0.3004	0.3005	0.3005



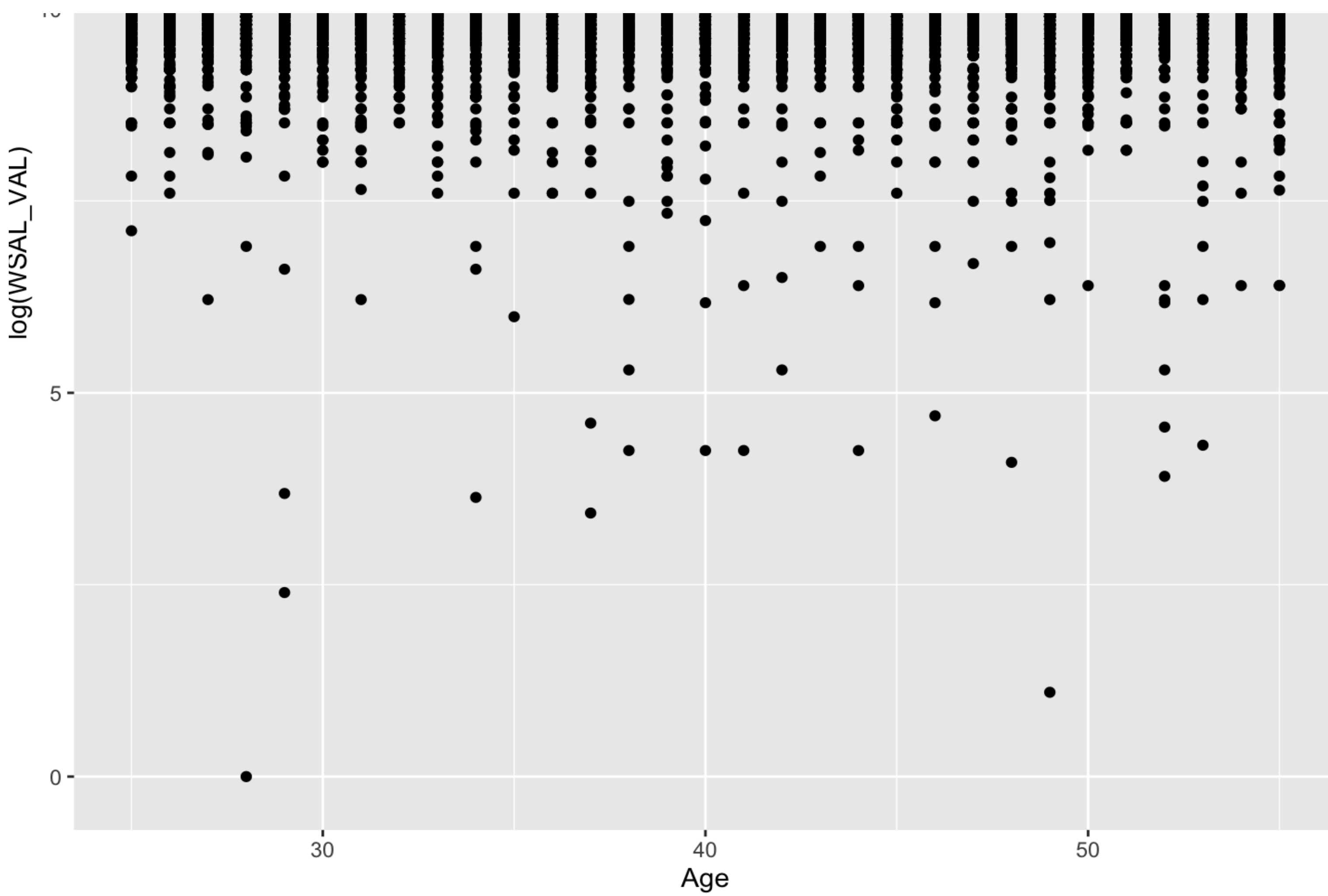


Table above show four different regression model to predict the percent change in wages. There are linear, quadratic, cubic and quartic. Based on the graph, the best fit model will be quadratic model. Also, by looking at the Age, Age^2, Age^3, and Age^4 coefficients at those four model, Age^3 or Age^4 is not statisitical significiant at 0.1% level. (highlight as red) I also tested the when age^2, age^3, age^4 are equal to 0, the test showed that they are significant at 0.1% level. (highlight as blue above) As people getting older, they will pay more at a nonconstant rate. Also, they will reach the highest salary at their position at sometime. Eventually, they will retired and getting pay so little.(negative non constant rate)

Question 3 - Nonlinear Regression Model of Log Wages and Salary

Regressor	(1)	(2)	(3)	(4)
Age	0.0240531*** (0.0004239)	1.453e-01*** (1.445e-03)	3.180e-01*** (5.342e-03)	6.598e-01*** (1.787e-02)
Age^2		-1.463e-03*** (1.705e-05)	-5.432e-03*** (1.280e-04)	-1.749e-02*** (6.560e-04)
Age^2*married		-7.237e-05*** (6.154e-06)	-2.270e-04*** (3.704e-05)	-1.266e-03*** (1.646e-04)
Age^3			2.787e-05*** (9.586e-07)	2.023e-04*** (1.002e-05)
Age^3*married			2.841e-06*** (4.854e-07)	3.460e-05*** (4.381e-06)
Age^4				-8.856e-07*** (5.410e-08)

Age^4*married				-2.543e-07*** (3.150e-08)
female	-0.4697657*** (0.0067713)	-4.497e-01*** (6.501e-03)	-4.431e-01*** (6.454e-03)	-4.418e-01*** (6.430e-03)
Afam(African)	-0.0115788 (0.0109154)	-6.055e-02*** (1.049e-02)	-7.530e-02*** (1.042e-02)	-7.907e-02*** (1.038e-02)
Asian	0.0543877*** (0.0157622)	5.932e-02*** (1.512e-02)	5.899e-02*** (1.501e-02)	5.312e-02*** (1.495e-02)
Amindian	-0.0940318** (0.0302812)	-1.245e-01*** (2.906e-02)	-1.386e-01*** (2.884e-02)	-1.453e-01*** (2.873e-02)
race_oth	-0.0511596* (0.0214743)	-5.756e-02** (2.060e-02)	-6.151e-02** (2.045e-02)	-6.428e-02** (2.037e-02)
Hispanic	0.0678878*** (0.0109940)	2.646e-02* (1.056e-02)	3.737e-03 (1.050e-02)	-1.144e-02 (1.047e-02)
educ_hs	0.6894889*** (0.0125442)	5.429e-01*** (1.214e-02)	4.747e-01*** (1.218e-02)	4.198e-01*** (1.232e-02)
educ_smcoll	0.7281593*** (0.0131749)	6.067e-01*** (1.271e-02)	5.460e-01*** (1.272e-02)	4.900e-01*** (1.286e-02)
educ_as	0.9753892*** (0.0150049)	7.854e-01*** (1.455e-02)	7.051e-01*** (1.459e-02)	6.447e-01*** (1.472e-02)
educ_bach	1.3011988*** (0.0132458)	1.109e+00*** (1.288e-02)	1.021e+00*** (1.300e-02)	9.586e-01*** (1.317e-02)
educ_adv	1.5801872*** (0.0147787)	1.397e+00*** (1.432e-02)	1.310e+00*** (1.440e-02)	1.254e+00*** (1.452e-02)
married	1.3844145*** (0.0225149)	3.452e-01*** (1.350e-02)	3.356e-01*** (2.701e-02)	5.944e-01*** (5.879e-02)

dividsep	0.1610399*** (0.0132270)	7.384e-02*** (1.264e-02)	5.712e-02*** (1.272e-02)	7.179e-02*** (1.271e-02)
union_m	0.2529393*** (0.0240974)	2.120e-01*** (2.313e-02)	2.154e-01*** (2.295e-02)	2.166e-01*** (2.286e-02)
veteran	-0.0734519*** (0.0140374)	1.194e-02 (1.350e-02)	-1.221e-02 (1.342e-02)	-9.952e-03 (1.337e-02)
immigrant	0.0077123*** (0.0142725)	-9.710e-02*** (1.374e-02)	-1.135e-01*** (1.365e-02)	-1.190e-01*** (1.360e-02)
immig2gen	-0.0127274 (0.0131068)	4.456e-02*** (1.259e-02)	3.959e-02** (1.250e-02)	4.018e-02** (1.245e-02)
Intercept	8.2895355*** (0.0169877)	6.309e+00*** (2.766e-02)	4.165e+00*** (6.690e-02)	9.093e-01*** (1.675e-01)
F-Statistics and p-Values on Joint Hypotheses				
age and all interactions = 0	1612 (< 2.2e-16) ***	3886.6 (< 2.2e-16) ***	2658.6 < 2.2e-16 ***	2016.8 < 2.2e-16 ***
F statistic for the whole model	2157 (< 2.2e-16)	2634 (< 2.2e-16)	2497 (< 2.2e-16)	2336 (< 2.2e- 16)
SER	1.007	0.9659	0.9586	0.955
R^2	0.3022	0.3575	0.3672	0.372

We have added age*married into previous models. According to the results, we have found all p-value in models are less than 0.05 (p-value: < 2.2e-16), so we reject the null hypothesis that β = 0. Hence there is a significant relationship between the variables in the linear regress model of data set “noZeroWage”.