# Final Project

Walker M Johnpaul C

2024-04-04

head

```
#install.packages("freqparcoord")
#install.packages("regclass")
library(freqparcoord)
```

```
## Loading required package: parallel
```

```
## Loading required package: ggplot2
```

```
## Loading required package: GGally
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
## Loading required package: FNN
```

```
## Loading required package: mvtnorm
```

```
##
##
##
##
##
##    For a quick introduction, type ?freqparcoord, and
##    run the examples, making sure to read the comments.
##
##
##
```

```
library(dplyr)
```

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

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

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

```
library(car)
```

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

```
##
```

```
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode
```

```
data(prgeng)
head(prgeng)
```

```
##        age cit educ engl occ birth sex wageinc wkswrkd yrentry powspuma
## 1 50.30082   1   13    0 102     6   2   75000      52       0     6010
## 2 41.10139   1    9    0 101     6   1   12300      20       0        0
## 3 24.67374   1    9    0 102     6   2   15400      52       0     6010
## 4 50.19951   1   11    0 100     8   1       0      52       0        0
## 5 51.18112   1   11    0 100     6   2     160       1       0     6010
## 6 57.70413   1   11    0 100     6   1       0       0       0        0
```

correlation matrix

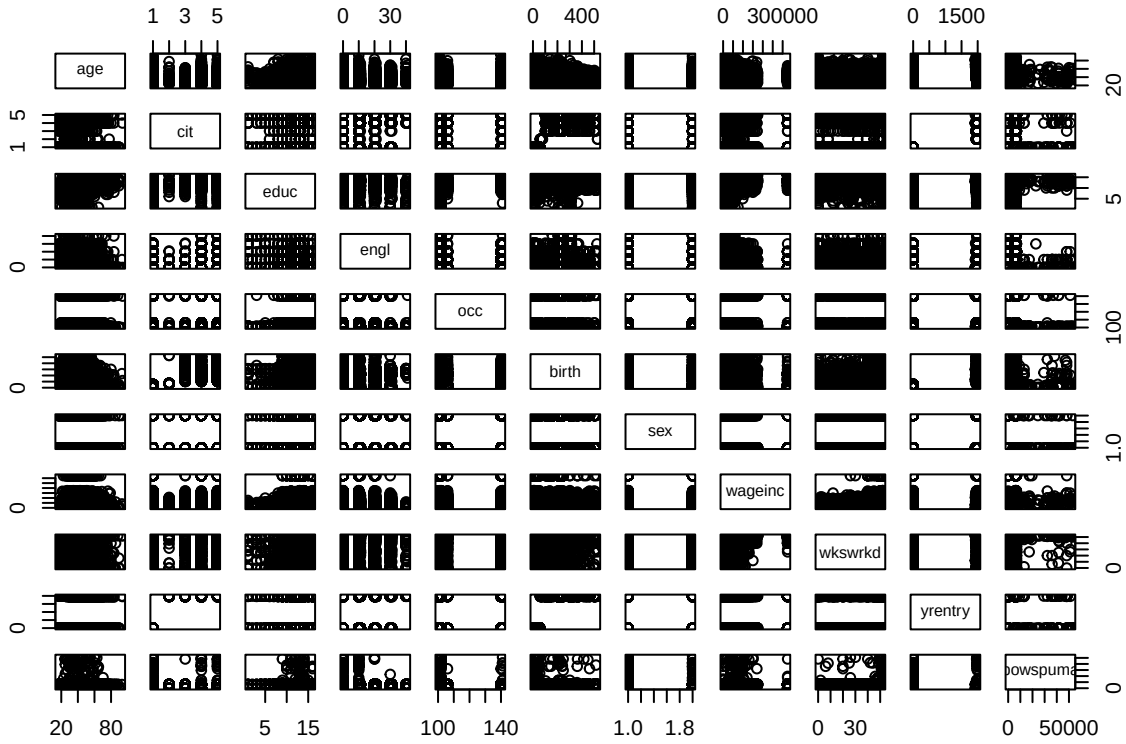
```
cor(prgeng)
```

```
##                  age          cit        educ        engl          occ
## age       1.000000000 -0.211662252  0.06195199 -0.15418225  0.114399273
## cit      -0.211662252  1.000000000  0.24397616  0.67435598  0.031129699
## educ      0.061951990  0.243976163  1.00000000  0.12732575  0.104156992
## engl     -0.154182254  0.674355983  0.12732575  1.00000000  0.036996559
## occ       0.114399273  0.031129699  0.10415699  0.03699656  1.000000000
## birth    -0.158221901  0.902685257  0.19723215  0.62508870  0.032958868
## sex      -0.004094500  0.014592088 -0.04779077  0.02642689 -0.147953742
## wageinc   0.113110192  0.033628497  0.23902018 -0.02589813  0.089887026
## wkswrkd  -0.003806349 -0.025809927  0.11449646 -0.04205652  0.032825140
## yrentry  -0.185313353  0.978709509  0.23540714  0.67626509  0.039204577
## powspuma -0.068537124  0.008699721  0.10094696 -0.01818481 -0.003951315
##                birth         sex     wageinc      wkswrkd     yrentry
## age      -0.15822190 -0.00409450  0.11311019 -0.003806349 -0.18531335
## cit       0.90268526  0.01459209  0.03362850 -0.025809927  0.97870951
## educ      0.19723215 -0.04779077  0.23902018  0.114496455  0.23540714
## engl      0.62508870  0.02642689 -0.02589813 -0.042056518  0.67626509
## occ       0.03295887 -0.14795374  0.08988703  0.032825140  0.03920458
## birth     1.00000000  0.01916462  0.03281126 -0.010511277  0.92501212
## sex       0.01916462  1.00000000 -0.11584422 -0.051377496  0.02130362
## wageinc   0.03281126 -0.11584422  1.00000000  0.415328556  0.04022630
## wkswrkd  -0.01051128 -0.05137750  0.41532856  1.000000000 -0.01289093
## yrentry   0.92501212  0.02130362  0.04022630 -0.012890926  1.00000000
## powspuma  0.01317393 -0.05042995  0.18175956  0.362563768  0.01044344
##              powspuma
## age      -0.068537124
## cit       0.008699721
## educ      0.100946956
## engl     -0.018184813
## occ      -0.003951315
## birth     0.013173932
## sex      -0.050429950
## wageinc   0.181759559
## wkswrkd   0.362563768
## yrentry   0.010443438
```

```
## powspuma  1.000000000
```

High correlations between cit, birth, and yrentry

```
pairs(prgeng)
```




initial model with all variables

```
bigmodel<-lm(wageinc~age+ educ+engl+occ+birth+factor(sex)+wkswrkd+yrentry+powspuma+cit, data =prgeng)
sumBig<-summary(bigmodel)
sumBig
```

```
##
## Call:
## lm(formula = wageinc ~ age + educ + engl + occ + birth + factor(sex) +
##     wkswrkd + yrentry + powspuma + cit, data = prgeng)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -101536  -20265   -4606   12684  290324
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -9.240e+04  3.387e+03 -27.282  < 2e-16 ***
## age           4.570e+02  2.818e+01  16.220  < 2e-16 ***
## educ          5.074e+03  1.947e+02  26.062  < 2e-16 ***
## engl         -3.779e+02  5.616e+01  -6.729 1.75e-11 ***
## occ           1.130e+02  2.018e+01   5.600 2.17e-08 ***
## birth        -8.887e+00  7.373e+00  -1.205   0.2281
## factor(sex)2 -9.178e+03  7.125e+02 -12.881  < 2e-16 ***
## wkswrkd       1.266e+03  2.235e+01  56.661  < 2e-16 ***
## yrentry       3.404e+00  1.738e+00   1.958   0.0502 .
```

```
## powspuma      4.578e-01  1.062e-01   4.310 1.64e-05 ***
## cit           3.894e+02  8.665e+02   0.449   0.6531
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 42760 on 20079 degrees of freedom
## Multiple R-squared:  0.2319, Adjusted R-squared:  0.2315
## F-statistic: 606.2 on 10 and 20079 DF,  p-value: < 2.2e-16
```

```
anova(bigmodel)
```

```
## Analysis of Variance Table
##
## Response: wageinc
##                Df     Sum Sq    Mean Sq  F value    Pr(>F)
## age             1 6.1150e+11 6.1150e+11  334.440 < 2.2e-16 ***
## educ            1 2.5828e+12 2.5828e+12 1412.565 < 2.2e-16 ***
## engl            1 8.1232e+10 8.1232e+10   44.427 2.709e-11 ***
## occ             1 1.5406e+11 1.5406e+11   84.258 < 2.2e-16 ***
## birth           1 5.5798e+10 5.5798e+10   30.517 3.351e-08 ***
## factor(sex)     1 4.4330e+11 4.4330e+11  242.450 < 2.2e-16 ***
## wkswrkd         1 7.0820e+12 7.0820e+12 3873.278 < 2.2e-16 ***
## yrentry         1 3.8099e+10 3.8099e+10   20.837 5.030e-06 ***
## powspuma        1 3.3961e+10 3.3961e+10   18.574 1.642e-05 ***
## cit             1 3.6937e+08 3.6937e+08    0.202    0.6531
## Residuals   20079 3.6713e+13 1.8284e+09
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

vif

```
vifs<-vif(bigmodel)
vifs
```

```
##         age        educ        engl         occ       birth factor(sex)
##    1.098723    1.118299    1.865848    1.049403    6.968091    1.029858
##     wkswrkd     yrentry    powspuma         cit
##    1.170107   31.148294    1.165435   24.825996
```

These three have multicollinearity lets kick them and test

```
high_vif_vars <- names(vifs[vifs > 10])
high_vif_vars
```

```
## [1] "yrentry" "cit"
```

get rid of yrentry and cit

```
reducedBirth<-lm(wageinc~age+ educ+engl+occ+factor(sex)+wkswrkd+powspuma+birth, data =prgeng)
summary(reducedBirth)
```

```
##
## Call:
## lm(formula = wageinc ~ age + educ + engl + occ + factor(sex) +
##     wkswrkd + powspuma + birth, data = prgeng)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -101606  -20263   -4593   12697  289181
```

```
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -9.352e+04  3.235e+03 -28.912  < 2e-16 ***
## age           4.418e+02  2.771e+01  15.944  < 2e-16 ***
## educ          5.223e+03  1.914e+02  27.287  < 2e-16 ***
## engl         -2.917e+02  5.295e+01  -5.509 3.66e-08 ***
## occ           1.142e+02  2.018e+01   5.660 1.53e-08 ***
## factor(sex)2 -9.146e+03  7.124e+02 -12.838  < 2e-16 ***
## wkswrkd       1.265e+03  2.231e+01  56.708  < 2e-16 ***
## powspuma      4.520e-01  1.063e-01   4.253 2.12e-05 ***
## birth         2.063e+01  3.639e+00   5.669 1.46e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 42780 on 20081 degrees of freedom
## Multiple R-squared:  0.2311, Adjusted R-squared:  0.2308
## F-statistic: 754.3 on 8 and 20081 DF,  p-value: < 2.2e-16
```

```
vifreducedBirth<-vif(reducedBirth)
vifreducedBirth
```

```
##         age        educ        engl         occ factor(sex)     wkswrkd
##    1.061674    1.080204    1.657076    1.048210    1.028410    1.164268
##    powspuma       birth
##    1.165269    1.695737
```

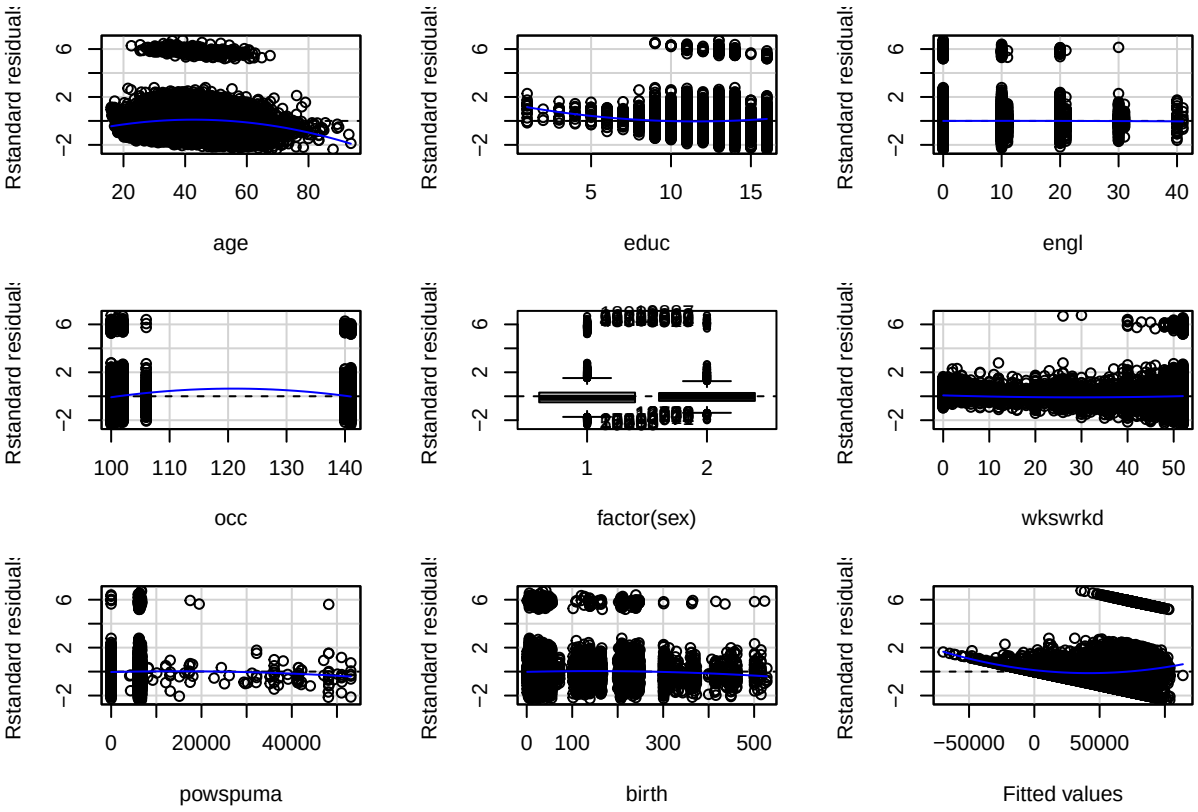
Looking at different combos and individual predictors, birth seems to be the best of 3 multicollinear variables to keep.

```
f_test_result <- anova(reducedBirth,bigmodel)
f_test_result
```

```
## Analysis of Variance Table
##
## Model 1: wageinc ~ age + educ + engl + occ + factor(sex) + wkswrkd + powspuma +
##     birth
## Model 2: wageinc ~ age + educ + engl + occ + birth + factor(sex) + wkswrkd +
##     yrentry + powspuma + cit
##   Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
## 1  20081 3.6752e+13
## 2  20079 3.6713e+13  2 3.9326e+10 10.754 2.148e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

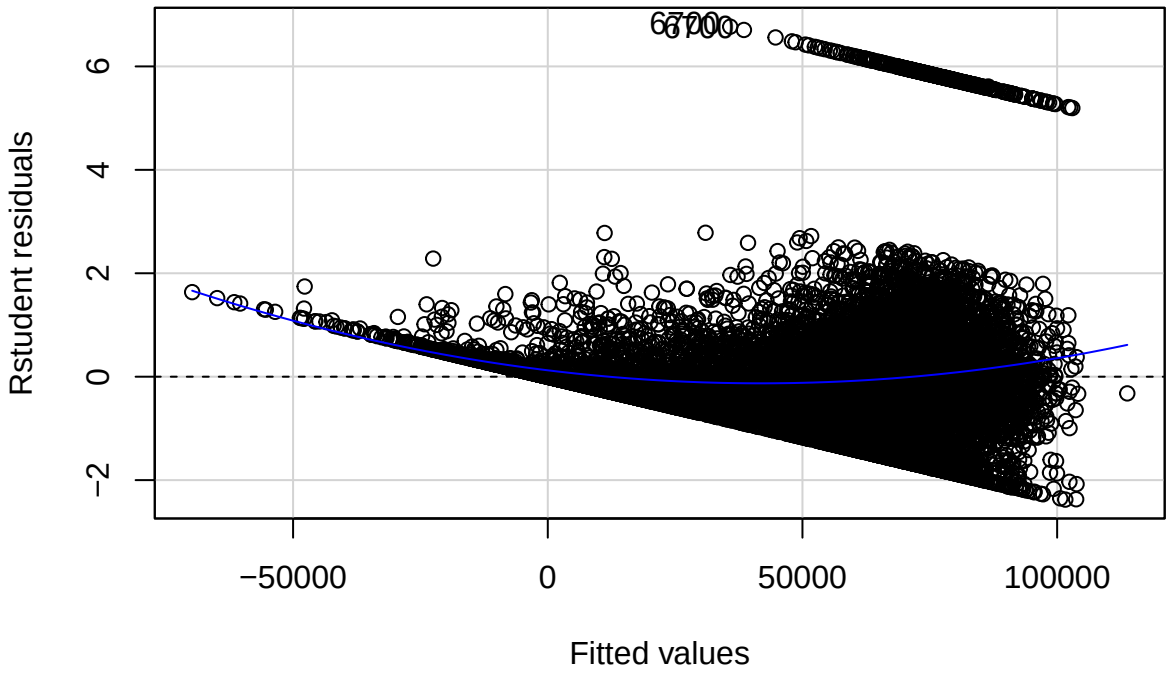
residual plots

```
residualPlots(reducedBirth, type="rstandard")
```




```
##              Test stat Pr(>|Test stat|)
## age           -19.4182       < 2.2e-16 ***
## educ            9.1971       < 2.2e-16 ***
## engl           -0.4526        0.650843
## occ            -9.4955       < 2.2e-16 ***
## factor(sex)
## wkswrkd         4.7381       2.172e-06 ***
## powspuma       -3.1647        0.001555 **
## birth          -5.3993       6.766e-08 ***
## Tukey test     17.2080       < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

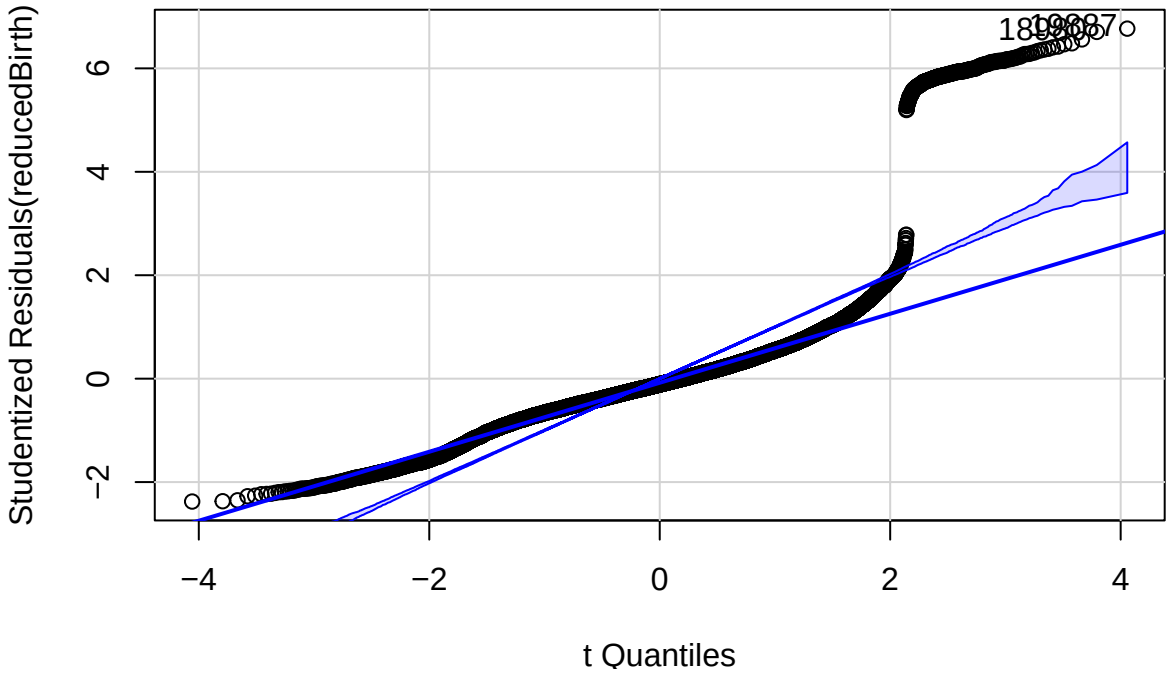
```
residualPlots(reducedBirth, ~1, type="rstudent",
id=list(labels=prgeng$powspuma))
```




```
##            Test stat Pr(>|Test stat|)
## Tukey test    17.208        < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

qqplot

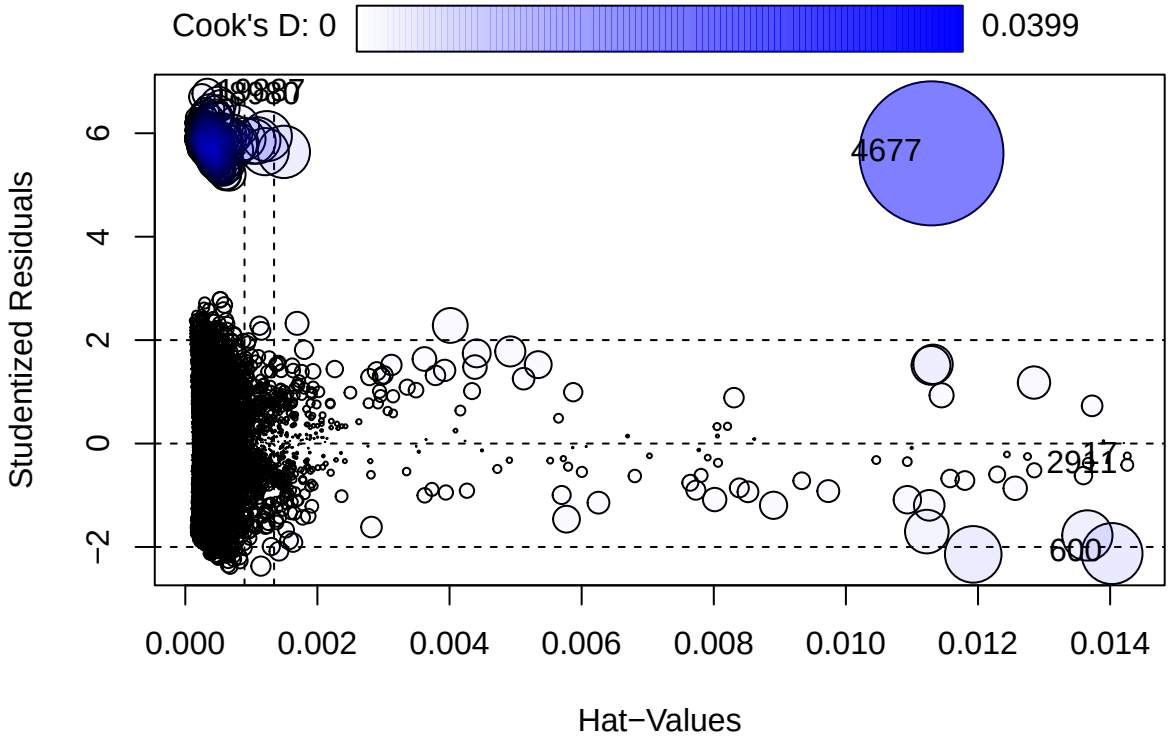
```
qqPlot(reducedBirth)
```




```
## [1] 18980 19887
```

cook's d

```
influencePlot(reducedBirth, id=list(labels=row.names(prgeng)))
```




```
##          StudRes          Hat        CookD
## 17    -0.2399878 0.0142596081 9.257663e-05
## 600   -2.1275633 0.0140286975 7.154830e-03
## 2911  -0.4124774 0.0142596939 2.734786e-04
## 4677   5.6114064 0.0112950726 3.990841e-02
## 18980  6.7049692 0.0002506309 1.249525e-03
## 19887  6.7682334 0.0003317605 1.685423e-03
```

```
ind <- prgeng %>%
  filter(row.names(prgeng) == "4677")
ind
```

```
##          age cit educ engl occ birth sex wageinc wkswrkd yrentry powspuma
## 4677 46.8542   4   11    0 100   139   1  325000      52    1990    48180
```

```
ind <- prgeng %>%
  filter(wageinc > 300000)
ind
```

```
##           age cit educ engl occ birth sex wageinc wkswrkd yrentry powspuma
## 66    32.95433   1    9    0 101    28   1  325000      52       0     6010
## 172   55.28573   1   14    0 100    11   1  325000      52       0     6030
## 214   35.53378   1   11    0 100    48   1  325000      52       0     6030
## 454   54.67283   1   13    0 101     6   1  325000      52       0     6040
## 484   36.21500   1   13    0 102    21   1  325000      52       0     6040
## 489   44.19617   1   13    0 101     6   1  325000      40       0     6040
## 499   42.79265   1   13    0 102     6   1  325000      52       0     6120
## 522   41.14038   4   16   10 101   217   1  325000      40    1970     6040
## 649   34.82870   5   14   11 100   109   1  325000      52    1995     6040
## 730   39.68378   1   14    0 102    36   1  325000      52       0     6140
```

```
## 740   40.89209   1   13    0 101     6   1  325000       52        0      6120
## 742   39.46099   1   14    0 102    48   1  325000       52        0      6130
## 761   58.98007   1   14    0 102    20   1  325000       52        0      6130
## 815   52.54049   1   14    0 100    34   1  325000       52        0      6050
## 818   39.72499   1   11    0 100    21   1  325000       52        0      6050
## 1373  43.82735   1   13    0 100    42   1  325000       52        0      6040
## 1964  50.36843   4   11    0 100   209   1  325000       52     1959      6309
## 1979  46.75900   1   14    0 102     6   1  325000       52        0      6060
## 2019  37.26979   1   14    0 100    36   2  325000       40        0      6070
## 2029  55.38466   1   14    0 140    49   1  325000       52        0      6070
## 2068  55.78213   1   11    0 141    17   1  325000       52        0      6090
## 2438  34.70363   4   13   10 102   231   1  325000       52     1984      6120
## 2494  53.72045   4   12    0 100   226   1  325000       52     1978      6150
## 2615  39.99202   1   13    0 102     6   1  325000       52        0      6160
## 2625  38.39620   1   14    0 100     6   1  325000       52        0      6130
## 2652  54.36779   1   13    0 141    42   1  325000       52        0      6140
## 2689  51.90092   1   14    0 141    41   1  325000       52        0      6120
## 2697  46.50407   1   16    0 100    42   1  325000       52        0      6130
## 2747  32.19572   1   13    0 100    15   1  325000       52        0      6120
## 2776  55.92571   1   13    0 100    20   2  325000       52        0         0
## 2789  36.22151   1   14    0 102    39   1  325000       51        0      6130
## 2801  34.92126   1   13    0 102     6   1  325000       52        0      6160
## 2851  52.05397   4   16   10 102   231   1  325000       52     1981      6150
## 2859  59.21782   1   12    0 100    36   2  325000       52        0      6120
## 2891  48.41932   1   14    0 102    39   1  325000       52        0      6160
## 2892  51.55972   1    9    0 102    29   2  325000       52        0      6160
## 3126  42.14737   1   11    0 101    13   1  325000       52        0      6130
## 3138  34.50314   1   13    0 102     6   1  325000       52        0      6160
## 3218  47.18820   1   14    0 100    36   1  325000       52        0      6130
## 3403  27.23804   1   14    0 100    51   2  325000       52        0      6130
## 3448  45.31565   1   11    0 100    53   1  325000       50        0      6130
## 3552  42.79832   1   13    0 100    26   2  325000       52        0      6130
## 3558  43.23799   5   11    0 101   139   1  325000       52     1980      6130
## 3659  28.53811   5   14    0 102   210   1  325000       52     1993      6130
## 3690  47.77261   1   13    0 100    15   1  325000       52        0      6130
## 3740  47.46625   1   13    0 101    18   1  325000       52        0      6140
## 3812  52.10545   1   13    0 102     6   2  325000       52        0      6130
## 3924  42.04694   1   12    0 100     6   2  325000       52        0      6140
## 3972  25.44962   1   11    0 102     6   1  325000       50        0      6160
## 4083  46.41838   4   13    0 100   436   1  325000       52     1957      6140
## 4086  30.94949   1   14    0 100    55   2  325000       52        0      6130
## 4099  32.05816   5   13    0 101   138   2  325000       52     1992      6130
## 4108  54.11874   1   14    0 100    42   2  325000       52        0      6140
## 4203  35.07530   1   13    0 102     6   1  325000       52        0      6140
## 4209  31.49432   4   13    0 102   209   1  325000       52     1974      6140
## 4213  52.52353   1   12    0 102     6   1  325000       52        0      6140
## 4226  48.49693   4   14   10 100   240   2  325000       50     1974         0
## 4257  44.54085   1   13    0 102    36   1  325000       52        0      6140
## 4356  53.09867   5   13   10 101   363   1  325000       52     1985      6160
## 4358  42.40473   1   16    0 102    46   1  325000       52        0      6160
## 4398  33.02003   5   14   10 102   109   1  325000       52     1993      6140
## 4442  36.66846   5   14   10 102   243   1  325000       52     1981      6160
## 4445  53.66012   1   16    0 141    41   1  325000       52        0      6150
## 4446  42.01863   1   16    0 100    36   1  325000       52        0      6160
```

```
## 4453  45.32955   1   12    0 100     6   1  325000       52        0      6140
## 4485  50.20606   1   16    0 102    25   1  325000       52        0      6140
## 4506  30.67428   1   13   10 100    34   1  325000       52        0      6130
## 4525  25.87300   5   14   10 102   163   1  325000       52     1998      6160
## 4550  61.68453   1   16    0 100    31   1  325000       52        0      6160
## 4555  44.20927   1   13    0 102    53   1  325000       52        0      6160
## 4616  39.75867   1   11    0 140    25   1  325000       50        0      6150
## 4619  51.99249   1   16    0 140     6   1  325000       52        0      6160
## 4634  35.13831   5   14   10 141   215   1  325000       50     1995      6160
## 4656  52.15498   1   14    0 102     6   1  325000       50        0      6160
## 4667  40.20777   5   16    0 140   210   1  325000       52     1985      6150
## 4675  40.04842   1   16    0 102     6   1  325000       52        0      6160
## 4677  46.85420   4   11    0 100   139   1  325000       52     1990     48180
## 4685  36.06261   1   10    0 140     6   1  325000       50        0      6130
## 4687  43.95885   1   14    0 100    24   1  325000       52        0      6160
## 4711  44.91458   1   13    0 102    36   1  325000       52        0      6160
## 4723  38.17702   5   14   10 102   109   1  325000       52     1996      6140
## 4725  45.43360   5   13   10 102   110   1  325000       52     1988      6160
## 4736  62.32631   4   16    0 100   102   1  325000       52     1938      6160
## 4746  50.73392   5   13    0 102   138   1  325000       48     1962      6150
## 4862  22.55638   5   13   10 100   211   2  325000       52     1993      6130
## 4974  28.93919   1   13    0 100     6   1  325000       52        0      6160
## 5048  37.46442   1   13    0 100    48   1  325000       52        0      6140
## 5055  49.39699   1   13    0 102    17   1  325000       52        0      6160
## 5159  37.11748   1   14    0 102    18   1  325000       52        0      6150
## 5223  57.79332   1   14    0 101    20   1  325000       52        0      6150
## 5295  42.24907   1   13    0 141     6   2  325000       50        0      6150
## 5344  43.14828   4   14   20 141   240   1  325000       50     1973      6160
## 5446  47.95688   4   14   10 101   163   1  325000       50     1978      6120
## 5458  40.55066   4   13   20 102   209   1  325000       50     1977      6150
## 5509  32.33135   4   13   10 102   207   2  325000       52     1985      6150
## 5694  34.92062   5   14   10 141   210   1  325000       52     1986      6160
## 5710  38.27182   1   14    0 140    27   1  325000       52        0      6140
## 5761  41.26838   4   16   20 102   209   1  325000       52     1985      6160
## 5864  35.86416   1   13    0 102     6   1  325000       40        0      6160
## 5897  33.07197   5   14   10 141   210   1  325000       50     1990      6160
## 5904  38.46776   4   13   10 140   226   1  325000       52     1982      6040
## 6057  46.38047   1   13    0 140     6   1  325000       52        0      6160
## 6166  34.40380   4   14   10 141   210   1  325000       52     1971      6160
## 6208  46.79199   4   13   20 140   247   1  325000       52     1982      6160
## 6215  41.65616   5   14   10 102   210   1  325000       52     1991      6160
## 6256  33.99182   5   14   10 102   240   1  325000       52     1989      6160
## 6311  41.99323   1   13    0 100     6   2  325000       51        0      6160
## 6312  51.30405   1   12    0 141     6   1  325000       52        0      6160
## 6322  28.46205   5   14   10 101   134   1  325000       51     1998      6150
## 6329  38.07461   1   13    0 102    29   1  325000       52        0      6160
## 6399  30.12733   5   13   10 102   501   1  325000       52     2000      6150
## 6471  29.09175   1   14    0 101    34   1  325000       49        0      6150
## 6498  30.40312   1   13    0 102     6   1  325000       50        0      6140
## 6541  47.76379   1   16    0 100    26   1  325000       52        0      6140
## 6552  36.94811   1   13    0 102    25   2  325000       52        0      6160
## 6576  30.40418   4   11    0 101   129   1  325000       52     1982      6140
## 6616  37.88430   1   15    0 102    27   1  325000       52        0      6160
## 6680  52.37482   1   14    0 141    12   1  325000       52        0      6160
```

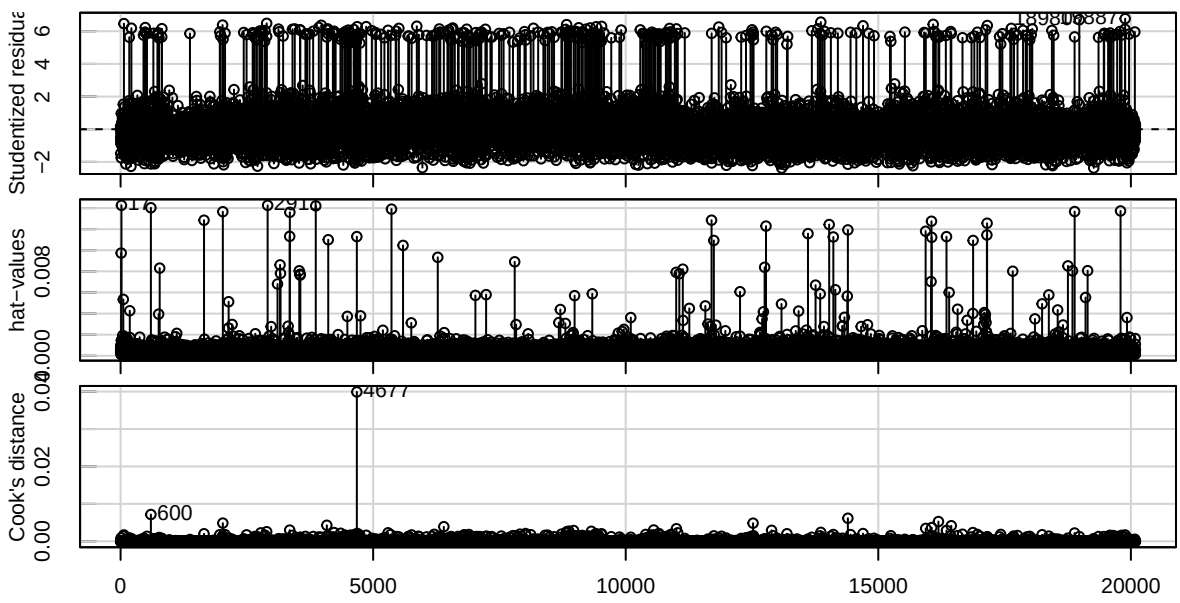
```
## 6693  49.34866   1   14    0 102    53   1  325000       52        0      6160
## 6760  48.89306   1   14    0 102     9   2  325000       52        0      6160
## 6771  55.66019   1   16    0 102    17   1  325000       52        0      6160
## 6801  36.37373   5   14   20 102   109   1  325000       52     1986      6160
## 6852  39.01314   1   13    0 102    29   2  325000       52        0      6160
## 6886  44.18022   1   13    0 141    34   1  325000       52        0      6160
## 6895  64.84395   1   16    0 100    25   1  325000       52        0      6160
## 6929  34.09674   1   16    0 100    25   1  325000       52        0      6160
## 6983  44.12598   1   16    0 100    45   1  325000       52        0      6160
## 7030  42.02623   5   16    0 102   139   1  325000       52     1979      6160
## 7031  39.29334   1   13    0 102     9   1  325000       52        0      6160
## 7043  51.30564   1   16    0 101    41   1  325000       52        0      6160
## 7099  47.31335   5   14    0 141   126   1  325000       52     1983      6160
## 7107  42.71195   1   13    0 140     6   1  325000       52        0      6160
## 7115  41.39062   1   13    0 101     6   1  325000       52        0      6150
## 7125  45.34879   1   14    0 140    36   1  325000       52        0      6160
## 7141  34.41227   1   13    0 102    36   1  325000       52        0      6160
## 7222  32.53073   5   13   20 100   210   1  325000       49     1997      6160
## 7250  46.13841   4   14   20 102   247   1  325000       52     1972      6160
## 7322  40.92010   4   14   10 140   212   1  325000       52     1976      6160
## 7348  44.19315   1   14    0 141     6   1  325000       52        0      6160
## 7493  41.88412   1   13    0 102    55   1  325000       52        0      6160
## 7538  46.90730   4   14   20 101   240   1  325000       52     1982      6140
## 7582  31.79136   5   14   10 102   210   1  325000       52     1990      6160
## 7700  44.12161   1   13   10 102     4   1  325000       52        0      6160
## 7768  46.97462   4   16   10 141   210   1  325000       52     1978      6160
## 7865  51.69794   1   16    0 100    35   1  325000       52        0      6160
## 7904  51.19696   5   14   20 141   243   1  325000       52     1982      6160
## 7996  32.54988   1   13    0 101    33   1  325000       52        0      6160
## 8014  30.29282   1   14    0 141     6   1  325000       52        0      6160
## 8017  67.54107   4   13    0 141   139   1  325000       52     1965      6160
## 8120  38.60562   4   14   20 141   247   1  325000       52     1980      6160
## 8173  34.02669   1   13    0 101     6   1  325000       45        0      6170
## 8267  33.65637   5   13   10 102   210   1  325000       52     1992      6150
## 8370  33.21092   5   14   10 102   210   1  325000       52     1990      6150
## 8404  36.74328   5   16   10 140   240   1  325000       52     1988      6160
## 8455  28.67707   1   13    0 140     1   1  325000       52        0      6160
## 8489  34.42639   4   13    0 101   209   1  325000       52     1996      6160
## 8539  32.83661   1   13    0 102    49   1  325000       52        0      6160
## 8581  42.62369   4   14   10 141   212   1  325000       52     1978      6160
## 8727  35.87403   4   13   20 141   209   1  325000       50     1976      6160
## 8742  33.38491   4   13   10 102   231   1  325000       52     1985      6160
## 8795  36.28199   4   14   10 102   209   1  325000       51     1982      6160
## 8807  35.76920   5   13   20 102   210   1  325000       48     1990      6160
## 8824  38.82565   5   14   10 102   210   1  325000       48     1989      6160
## 8830  48.16209   4   13   10 106   240   2  325000       40     1978         0
## 8833  29.07130   4   13   10 140   247   1  325000       52     1982         0
## 8836  38.90642   4   14   10 140   210   1  325000       52     1983      6160
## 8851  52.15650   4   12   20 141   209   1  325000       52     1979      6160
## 8875  40.46150   4   11    0 141   247   1  325000       52     1975      6160
## 8891  35.03184   4   14   10 100   247   1  325000       51     1979      6160
## 8903  38.64996   4   13   20 141   247   1  325000       52     1981      6160
## 8905  35.37998   4   14   10 101   202   1  325000       52     1989      6160
## 8914  33.00208   1   13    0 101    20   1  325000       51        0      6160
```

```
## 8968  30.87756   4   14   20 141   247   1  325000       48     1986        0
## 9003  56.65594   1   13    0 141    42   1  325000       52        0     6160
## 9004  43.60874   1   11    0 140    53   1  325000       52        0     6160
## 9046  34.89415   1   13    0 102    55   1  325000       52        0     6160
## 9070  44.20604   1   10    0 102    48   1  325000       52        0     6140
## 9088  48.44090   4   14   10 141   240   1  325000       52     1978     6160
## 9130  38.53627   4   14   10 102   207   1  325000       50     1982     6160
## 9139  48.98849   4   16   11 140   240   1  325000       52     1977     6160
## 9142  34.17673   1   13    0 102     6   1  325000       52        0     6160
## 9146  44.75855   1   13    0 100    41   1  325000       52        0     6160
## 9197  48.24978   4   14   10 140   240   1  325000       52     1980     6160
## 9213  29.83537   1   13    0 102    26   1  325000       48        0     6160
## 9229  46.55172   5   14   10 140   215   1  325000       52     1991     6160
## 9230  37.40909   1   14    0 100     6   1  325000       52        0     6160
## 9261  59.04878   1   14    0 141    36   1  325000       52        0     6160
## 9284  42.77731   4   14   10 102   212   1  325000       52     1976     6160
## 9299  36.18586   1   13    0 102    36   1  325000       43        0     6160
## 9320  39.08966   4   13    0 141   301   1  325000       52     1984     6160
## 9353  46.26001   1   13    0 102    13   1  325000       52        0     6140
## 9365  31.61607   3   13    0 102   217   1  325000       52     1990     6160
## 9382  36.04478   5   13    0 101   139   1  325000       52     1994     6160
## 9393  41.22221   4   14   10 102   209   1  325000       52     1979     6160
## 9407  49.28176   4   16   20 141   240   1  325000       52     1980     6160
## 9416  34.56858   5   10    0 102   127   1  325000       52     1994     6160
## 9429  42.14832   1   14    0 140    17   1  325000       52        0     6160
## 9432  45.72813   5   14   10 102   242   1  325000       52     1976     6160
## 9434  58.81037   4   14   10 141   233   1  325000       42     1968     6160
## 9460  49.01463   4   16   10 141   240   1  325000       52     1973     6160
## 9477  38.16607   5   14   10 102   210   1  325000       52     1993     6160
## 9514  50.82826   1   16   10 141    42   1  325000       52        0     6160
## 9523  42.44213   1   10    0 102    49   1  325000       52        0        0
## 9569  40.68797   1   13    0 102    26   1  325000       52        0     6160
## 9714  47.52128   1   14    0 141    26   1  325000       52        0     6160
## 9875  44.71797   4   14   10 141   163   1  325000       52     1989     6160
## 9910  44.79615   1   11    0 106    48   1  325000       52        0     6160
## 10288 33.55267   1   13    0 102    29   1  325000       48        0     6160
## 10335 44.37386   1   13    0 102     6   1  325000       52        0     6160
## 10356 47.92915   4   13   10 102   212   1  325000       50     1980     6150
## 10391 53.71024   1   13    0 102     6   1  325000       52        0     6160
## 10396 56.21684   5   16    0 141   166   1  325000       52     1995     6160
## 10420 53.91654   4   14   10 102   210   1  325000       50     1979     6140
## 10460 49.78396   1   13    0 140     6   1  325000       52        0     6160
## 10492 39.65189   1   14    0 102    31   1  325000       52        0     6160
## 10518 44.90394   5   13   20 102   120   1  325000       52     1995     6160
## 10554 60.83936   1   13   20 102    42   1  325000       52        0     6160
## 10559 34.07937   1   13    0 141     9   1  325000       52        0     6160
## 10589 40.39601   4   16   10 100   210   1  325000       52     1981     6150
## 10633 35.06972   5   14   10 102   210   1  325000       52     1987     6140
## 10655 33.45115   5   13    0 102   301   1  325000       52     1990     6160
## 10660 59.77516   1   14    0 101    36   1  325000       52        0     6170
## 10676 46.64814   1   16    0 141    25   1  325000       52        0     6160
## 10680 51.96290   1   14    0 101     6   1  325000       52        0     6160
## 10692 43.77697   5   13    0 102   139   1  325000       52     1983     6160
## 10795 45.14495   1   16    0 102     6   1  325000       52        0     6160
```

```
## 10802 36.46909   1   13    0 102    53   1  325000       52        0      6160
## 10807 42.03582   1   13    0 141     6   1  325000       52        0      6160
## 10815 42.63160   1   13    0 102    18   1  325000       52        0      6170
## 10859 39.66354   1   13    0 141    40   1  325000       52        0      6160
## 10868 38.33112   1   11    0 102     6   1  325000       52        0      6160
## 10871 50.88949   1   15    0 100     6   1  325000       52        0      6160
## 10885 35.30642   1   13    0 101     6   1  325000       52        0      6170
## 10886 50.18472   4   16   10 102   147   2  325000       52     1977      6160
## 10973 50.46883   1   13    0 141     6   1  325000       52        0      6170
## 11003 27.92157   4   13   30 102   209   1  325000       52     1973      6160
## 11030 43.80448   1   14    0 141     6   2  325000       52        0      6160
## 11169 53.03729   1   12    0 102     6   1  325000       52        0      6190
## 11701 44.30901   1   13    0 100     6   1  325000       52        0      6210
## 11842 30.18576   1   11    0 101     6   1  325000       52        0      6220
## 11897 49.95056   1   12    0 141    36   1  325000       48        0      6220
## 11988 51.72674   5   14    0 102   139   1  325000       52     1977      6220
## 12286 52.23078   1   14    0 100    53   1  325000       52        0      6500
## 12381 57.47369   1   13    0 100     6   1  325000       52        0      6230
## 12470 43.87178   4   13   10 141   214   1  325000       52     1975      6230
## 12502 41.02995   4   13   10 100   212   2  325000       52     1979      6309
## 12520 50.12874   4   13   10 101   247   1  325000       50     1975      6309
## 12521 55.15973   1   13    0 100     8   2  325000       49        0     17500
## 12523 47.90779   5   14   10 100   165   1  325000       50     1992      6309
## 12528 56.91161   4   14   10 141   210   1  325000       52     1969      6309
## 12783 30.93182   5   13   10 101   164   1  325000       52     1992      6309
## 12888 36.16600   4   13    0 102   363   1  325000       52     1970      6309
## 12917 57.25095   4   14   10 141   210   1  325000       52     1964      6160
## 12968 42.74008   4   13   10 101   158   1  325000       50     1977      6309
## 12998 47.25651   4   14   10 100   209   1  325000       48     1968      6700
## 13192 52.79976   4   16    0 141   217   1  325000       52     1954      6309
## 13206 47.98588   1   14    0 102    48   1  325000       52        0      6309
## 13685 35.28607   1   13   10 101     6   1  325000       52        0      6309
## 13804 41.09181   1   11    0 100     6   1  325000       52        0      6309
## 13846 34.51017   4   13    0 101   247   1  325000       52     1993      6309
## 13862 45.45507   1    9    0 101     6   2  325000       52        0      6309
## 13898 54.23221   1   13    0 100     6   1  325000       52        0      6309
## 13935 29.46686   4   13   10 101   210   1  325000       52     1982      6309
## 14061 47.67729   1   12    0 141     6   1  325000       50        0      6309
## 14200 54.77780   1   12    0 102     6   1  325000       52        0      6309
## 14451 60.66788   1   13    0 141    26   1  325000       40        0      6500
## 14565 35.95456   4   14   10 100   207   1  325000       49     1985      6309
## 14698 54.32740   1   10    0 100    36   2  325000       52        0      6309
## 14812 39.98425   5   13   10 102   231   1  325000       52     1986      6309
## 14916 55.74196   1   13    0 101    56   1  325000       52        0      6309
## 15231 50.48370   1   14    0 100     6   1  325000       52        0      6309
## 15248 42.91124   5   16   10 141   207   1  325000       52     1986      6500
## 15526 36.01437   1   13    0 100    27   1  325000       52        0      6309
## 15907 38.67221   1   13    0 102     6   1  325000       52        0      6309
## 15940 26.28835   4   14   21 100   523   1  325000       50     1999      6309
## 16087 34.90942   1   11    0 100    13   2  325000       52        0      6409
## 16100 45.43030   1   13    0 100     6   1  325000       49        0      6309
## 16106 31.97254   1   13    0 100    26   1  325000       52        0      6309
## 16139 49.29160   4   14   10 100   212   1  325000       52     1979      6309
## 16191 58.42801   1   14    0 101     6   1  325000       46        0     19500
```

```
## 16198 42.80237   1   14    0 106     6   1  325000       52        0     6309
## 16207 52.77733   1   14    0 100    17   1  325000       52        0     6309
## 16337 41.49300   4   14   10 102   212   1  325000       52     1984     6500
## 16433 37.72340   1   13    0 102    42   1  325000       52        0        0
## 16445 25.42323   3   13    0 102   416   1  325000       52     1975     6500
## 16667 51.14340   4   14   20 140   217   1  325000       52     1977     6500
## 16846 46.70842   1   14    0 140    27   1  325000       52        0     6500
## 16901 46.26287   4   14   10 100   209   1  325000       52     1977     6500
## 16962 32.95635   4   13   10 100   365   1  325000       50     1970     6500
## 17012 42.36841   4   14   10 141   240   1  325000       52     1984     6500
## 17125 34.73767   1   12    0 102     6   1  325000       52        0     6500
## 17158 33.92948   5   10   10 102   120   1  325000       52     1996     6500
## 17401 62.63234   1   13    0 101    26   1  325000       52        0     6409
## 17423 60.81740   1   16    0 141    26   1  325000       52        0     6309
## 17456 46.35197   1   13    0 100     6   1  325000       52        0     6309
## 17614 42.04928   1   14    0 100    56   1  325000       52        0     6500
## 17633 38.39726   1   11    0 100     6   1  325000       52        0     6140
## 17678 56.87534   1   14    0 141    17   1  325000       52        0     6500
## 17687 40.59969   4   13   10 101   212   1  325000       52     1979     6500
## 17756 32.70226   1   14    0 141     6   1  325000       52        0     6500
## 17866 35.87941   1   12    0 102    42   1  325000       52        0     6500
## 17918 44.52677   1   14    0 102     6   1  325000       52        0     6500
## 17942 43.76371   1   13    0 141    29   1  325000       52        0     6500
## 17998 51.10391   1   16    0 100    40   1  325000       50        0     6500
## 18028 50.62706   1   13    0 102    29   1  325000       52        0     6500
## 18048 39.61870   1   11    0 101    37   1  325000       52        0     6500
## 18437 28.56792   1   13    0 102    15   1  325000       51        0     6700
## 18472 42.63474   1   14    0 102    13   1  325000       52        0     6700
## 18886 59.69164   1   13    0 141    39   1  325000       52        0        0
## 18980 35.83790   1   13    0 102    35   1  325000       26        0     6700
## 19360 45.10797   1   14    0 141     6   1  325000       52        0     6700
## 19464 40.79084   1   13    0 141     6   1  325000       52        0     6700
## 19472 42.04571   1   13    0 102    26   1  325000       52        0     6700
## 19550 41.49316   1   13    0 141    18   1  325000       52        0     6700
## 19583 61.85109   1   13    0 141    24   1  325000       50        0     6700
## 19591 40.87292   1   13    0 102     6   1  325000       52        0     6700
## 19620 44.87442   5   13    0 141   138   1  325000       52     1997     6700
## 19655 30.99182   1   14    0 141    36   1  325000       52        0     6700
## 19729 37.25464   5   14   10 141   207   1  325000       48     1994     6700
## 19784 30.58714   1   13    0 102    47   1  325000       52        0     6700
## 19815 29.91037   1   14    0 141    39   1  325000       52        0     6700
## 19883 38.37176   1   13    0 102    30   2  325000       52        0     6700
## 19887 39.19821   1   13    0 100    42   2  325000       30        0     6700
## 19964 34.48807   1   14    0 102    42   1  325000       52        0     6700
## 20079 33.92606   1   13    0 101    36   1  325000       52        0     6700
```

```
influenceIndexPlot(reducedBirth, vars=c("Studentized","hat","Cook"), id=list(labels=row.names(prgeng)))
```




most influential

```
dfbeta.prgeng<-dfbetas(reducedBirth)

# Get the absolute values of the dfbetas for each predictor
abs_dfbeta.prgeng <- abs(dfbeta.prgeng)

# Find the index (case number) of the max value in each column
max_influence_case <- apply(abs_dfbeta.prgeng, 2, which.max)
```

As seen in the cooks disstance graph. This individual is influential, specifically within wkswrkd, powspuma

```
newprgeng <- prgeng %>%
  filter(row.names(prgeng) != "4677")
head(newprgeng)
```

```
##        age cit educ engl occ birth sex wageinc wkswrkd yrentry powspuma
## 1 50.30082   1   13    0 102     6   2   75000      52       0     6010
## 2 41.10139   1    9    0 101     6   1   12300      20       0        0
## 3 24.67374   1    9    0 102     6   2   15400      52       0     6010
## 4 50.19951   1   11    0 100     8   1       0      52       0        0
## 5 51.18112   1   11    0 100     6   2     160       1       0     6010
## 6 57.70413   1   11    0 100     6   1       0       0       0        0
```

```
reducedBirthNew<-lm(wageinc~age+ educ+engl+occ+factor(sex)+wkswrkd+powspuma+birth, data =newprgeng)
summary(reducedBirthNew)
```

```
##
## Call:
## lm(formula = wageinc ~ age + educ + engl + occ + factor(sex) +
##     wkswrkd + powspuma + birth, data = newprgeng)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -101569  -20242   -4560   12687  289316
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -9.352e+04  3.232e+03 -28.934  < 2e-16 ***
## age           4.397e+02  2.769e+01  15.877  < 2e-16 ***
## educ          5.240e+03  1.913e+02  27.393  < 2e-16 ***
## engl         -2.902e+02  5.291e+01  -5.485 4.19e-08 ***
## occ           1.145e+02  2.017e+01   5.678 1.38e-08 ***
## factor(sex)2 -9.141e+03  7.118e+02 -12.841  < 2e-16 ***
## wkswrkd       1.269e+03  2.230e+01  56.908  < 2e-16 ***
## powspuma      3.891e-01  1.068e-01   3.643  0.00027 ***
## birth         2.047e+01  3.636e+00   5.628 1.85e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 42750 on 20080 degrees of freedom
## Multiple R-squared:  0.2311, Adjusted R-squared:  0.2308
## F-statistic: 754.6 on 8 and 20080 DF,  p-value: < 2.2e-16
```

took it out. relationships look the same