Problem Set 3

MACS 30200 - Perspectives on Computational Research Luxi Han 10449918

Problem 1

a)

We first display the observations that are discrepant from the simple OLS model estimated. We display the observations with studendized residuals that is outside the range of [-2, 2]

```
## # A tibble: 82 × 9
##    biden female   age  educ   dem   rep     hat student  cooksd
##    <int>  <int> <int> <int> <int> <int>   <dbl>   <dbl>   <dbl>
## 1      0      1    70    12     0     1 0.00204   -2.91 0.00429
## 2      0      0    45    12     0     1 0.00142   -2.59 0.00237
## 3      0      0    40    14     0     0 0.00136   -2.50 0.00213
## 4     15      0    62     8     0     1 0.00411   -2.13 0.00466
## 5     15      1    20    13     0     0 0.00260   -2.12 0.00294
## 6      0      1    38    14     1     0 0.00122   -2.77 0.00233
## 7      0      0    34    12     0     0 0.00178   -2.57 0.00293
## 8      0      0    21    13     0     1 0.00259   -2.51 0.00407
## 9     15      1    29    12     0     1 0.00198   -2.18 0.00235
## 10     0      0    36    13     0     1 0.00149   -2.53 0.00239
## 11    15      1    86    12     0     0 0.00386   -2.28 0.00504
## 12    20      1    58     4     0     1 0.00922   -2.33 0.01262
## 13     0      0    56    11     0     0 0.00185   -2.65 0.00323
## 14     0      0    60    16     0     0 0.00236   -2.46 0.00358
## 15     0      1    28    12     1     0 0.00206   -2.83 0.00412
## 16     0      0    41    17     0     1 0.00252   -2.39 0.00360
## 17     0      1    90    16     0     1 0.00542   -2.79 0.01058
## 18     0      0    77    16     0     1 0.00394   -2.50 0.00615
## 19     0      1    51    16     0     1 0.00168   -2.72 0.00309
## 20     0      0    50    17     0     1 0.00257   -2.41 0.00372
## 21    15      1    81    16     0     1 0.00403   -2.12 0.00454
## 22     0      0    53    15     0     1 0.00161   -2.49 0.00249
## 23     8      1    52    12     1     0 0.00120   -2.52 0.00190
## 24     0      1    48    14     0     1 0.00104   -2.79 0.00200
## 25     0      0    64    12     0     1 0.00191   -2.62 0.00329
## 26     0      0    51    16     0     0 0.00198   -2.45 0.00296
## 27     0      1    31    16     0     1 0.00208   -2.68 0.00373
## 28    15      1    39    13     0     0 0.00119   -2.16 0.00138
## 29     0      0    46    13     0     1 0.00125   -2.55 0.00203
## 30    15      1    52    12     0     1 0.00120   -2.22 0.00147
## 31     5      0    51    16     0     1 0.00198   -2.23 0.00246
## 32    15      1    48    14     1     0 0.00104   -2.14 0.00118
## 33    15      1    36    14     0     0 0.00130   -2.11 0.00145
## 34     0      0    58    14     0     1 0.00155   -2.54 0.00249
## 35     0      1    23    12     0     0 0.00253   -2.82 0.00502
## 36     0      1    57    14     0     1 0.00121   -2.80 0.00237
## 37     0      0    70    12     0     1 0.00236   -2.64 0.00411
## 38    15      1    79    15     0     1 0.00327   -2.16 0.00381
```

```
## 39      0       0    35    13     0     0 0.00154    -2.53 0.00246
## 40      0       0    50    16     0     1 0.00196    -2.45 0.00293
## 41      0       0    78    16     0     0 0.00407    -2.50 0.00636
## 42      0       0    57    16     0     0 0.00220    -2.46 0.00332
## 43     15       1    42    17     0     0 0.00223    -2.01 0.00226
## 44      0       0    22    15     0     1 0.00260    -2.43 0.00384
## 45      0       0    78    12     0     1 0.00319    -2.65 0.00560
## 46      0       0    72     9     0     0 0.00392    -2.76 0.00745
## 47      0       0    62    14     0     1 0.00176    -2.54 0.00285
## 48     15       1    66    14     0     1 0.00170    -2.17 0.00200
## 49      0       1    91    14     0     1 0.00474    -2.87 0.00976
## 50     15       1    61    14     0     1 0.00139    -2.16 0.00162
## 51      0       0    50    14     0     0 0.00131    -2.52 0.00207
## 52      0       0    46    15     0     1 0.00150    -2.48 0.00230
## 53      0       0    54    17     0     1 0.00269    -2.41 0.00392
## 54      0       1    44    13     0     1 0.00105    -2.82 0.00209
## 55      0       1    58    12     0     0 0.00134    -2.88 0.00277
## 56      0       0    65    11     0     1 0.00227    -2.67 0.00402
## 57      0       0    63    17     0     0 0.00320    -2.43 0.00474
## 58     15       1    66    16     0     1 0.00241    -2.09 0.00264
## 59      0       1    34    14     0     1 0.00140    -2.76 0.00266
## 60      0       0    77    16     0     1 0.00394    -2.50 0.00615
## 61      0       0    62    14     0     1 0.00176    -2.54 0.00285
## 62     15       1    46    11     1     0 0.00156    -2.25 0.00197
## 63     15       1    48    14     0     1 0.00104    -2.14 0.00118
## 64     15       1    60    12     0     1 0.00141    -2.23 0.00176
## 65      0       0    39    12     0     0 0.00155    -2.58 0.00258
## 66      0       1    66    17     0     0 0.00305    -2.71 0.00558
## 67     15       1    41    14     0     1 0.00112    -2.12 0.00126
## 68     15       1    69    14     0     0 0.00193    -2.17 0.00229
## 69      0       0    32    16     1     0 0.00223    -2.41 0.00324
## 70      0       1    33    13     0     0 0.00148    -2.80 0.00289
## 71      0       1    24    15     0     0 0.00227    -2.71 0.00415
## 72      0       1    45    12     0     1 0.00122    -2.86 0.00248
## 73      0       0    27    14     0     0 0.00202    -2.48 0.00310
## 74      0       0    77    16     0     1 0.00394    -2.50 0.00615
## 75     15       1    57    17     0     0 0.00248    -2.04 0.00257
## 76     15       1    24    16     0     0 0.00260    -2.02 0.00264
## 77     15       1    65    15     0     0 0.00189    -2.13 0.00214
## 78     15       1    50    16     0     0 0.00166    -2.06 0.00177
## 79      0       1    62    14     0     0 0.00144    -2.81 0.00284
## 80      0       0    23    11     0     1 0.00303    -2.59 0.00508
## 81      0       0    70    12     1     0 0.00236    -2.64 0.00411
## 82     15       1    34    16     0     0 0.00192    -2.03 0.00199
```

Now we can use the cook distance to display the influential observations.

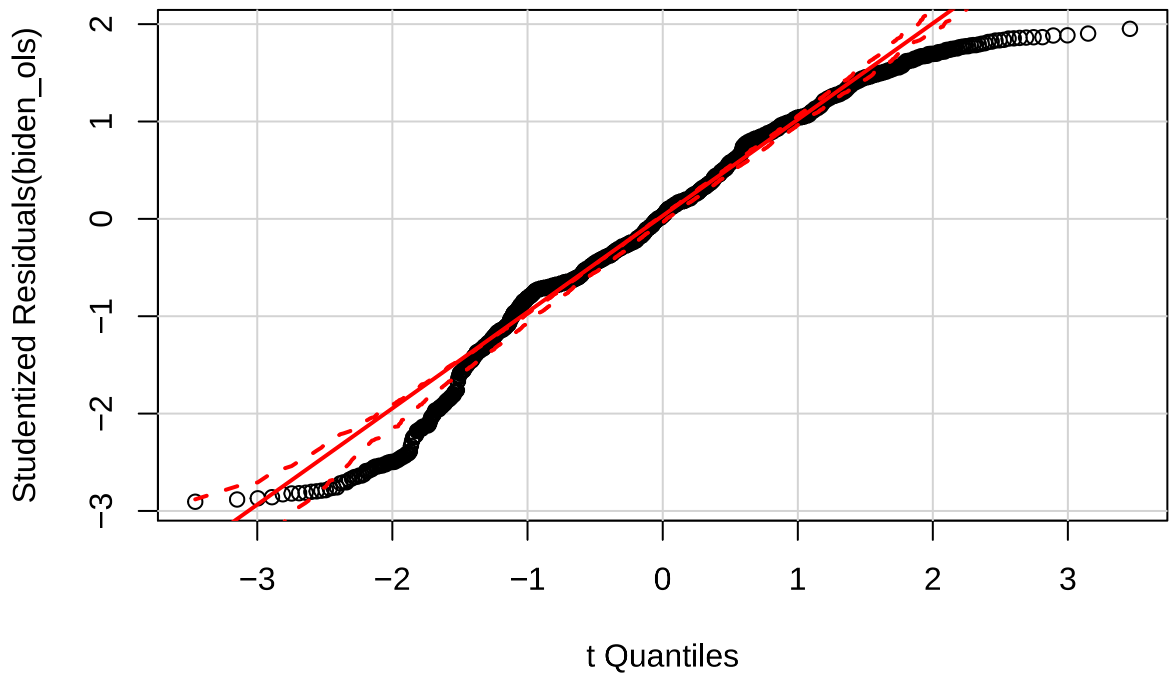
```
## # A tibble: 90 × 9
##    biden female   age  educ   dem   rep     hat student  cooksd
##    <int>  <int> <int> <int> <int> <int>   <dbl>   <dbl>   <dbl>
## 1      0      1    70    12     0     1 0.00204   -2.91 0.00429
## 2      0      0    45    12     0     1 0.00142   -2.59 0.00237
## 3     15      0    62     8     0     1 0.00411   -2.13 0.00466
## 4     15      1    20    13     0     0 0.00260   -2.12 0.00294
## 5    100      1    64     1     1     0 0.01537    1.02 0.00404
```

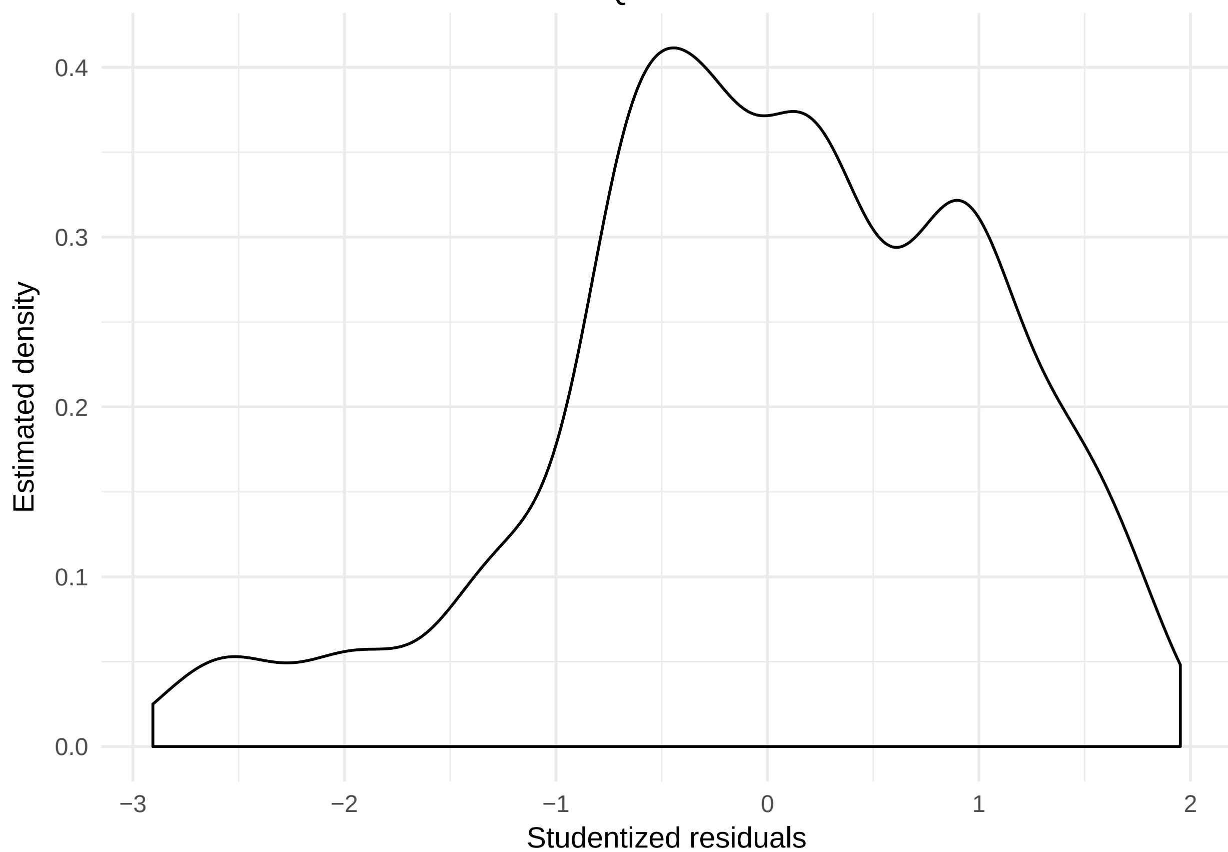
```
## 6     100       0     19     12      0      0 0.00304    1.78 0.00242
## 7     100       0     19     12      1      0 0.00304    1.78 0.00242
## 8       0       1     38     14      1      0 0.00122   -2.77 0.00233
## 9     100       1     76      3      1      0 0.01184    1.07 0.00344
## 10      0       0     34     12      0      0 0.00178   -2.57 0.00293
## 11      0       0     21     13      0      1 0.00259   -2.51 0.00407
## 12     15       1     29     12      0      1 0.00198   -2.18 0.00235
## 13      0       0     36     13      0      1 0.00149   -2.53 0.00239
## 14     15       1     86     12      0      0 0.00386   -2.28 0.00504
## 15     20       1     58      4      0      1 0.00922   -2.33 0.01262
## 16      0       0     56     11      0      0 0.00185   -2.65 0.00323
## 17    100       0     82      9      1      0 0.00497    1.56 0.00302
## 18      0       0     60     16      0      0 0.00236   -2.46 0.00358
## 19     30       0     40      5      0      0 0.00801   -1.55 0.00488
## 20      0       1     28     12      1      0 0.00206   -2.83 0.00412
## 21     15       0     22     12      0      1 0.00271   -1.90 0.00245
## 22      0       0     41     17      0      1 0.00252   -2.39 0.00360
## 23      0       1     90     16      0      1 0.00542   -2.79 0.01058
## 24      0       0     77     16      0      1 0.00394   -2.50 0.00615
## 25      0       1     51     16      0      1 0.00168   -2.72 0.00309
## 26      0       0     50     17      0      1 0.00257   -2.41 0.00372
## 27    100       1     78     17      1      0 0.00431    1.60 0.00278
## 28     15       1     81     16      0      1 0.00403   -2.12 0.00454
## 29      0       0     53     15      0      1 0.00161   -2.49 0.00249
## 30      0       0     64     12      0      1 0.00191   -2.62 0.00329
## 31      0       0     51     16      0      0 0.00198   -2.45 0.00296
## 32      0       1     31     16      0      1 0.00208   -2.68 0.00373
## 33      5       0     51     16      0      1 0.00198   -2.23 0.00246
## 34      0       0     58     14      0      1 0.00155   -2.54 0.00249
## 35     15       0     78     17      0      0 0.00476   -1.81 0.00391
## 36    100       0     82     12      1      0 0.00369    1.67 0.00259
## 37      0       1     23     12      0      0 0.00253   -2.82 0.00502
## 38     15       0     69     16      0      1 0.00306   -1.83 0.00256
## 39      0       1     57     14      0      1 0.00121   -2.80 0.00237
## 40     15       0     75     13      0      1 0.00278   -1.96 0.00266
## 41      0       0     70     12      0      1 0.00236   -2.64 0.00411
## 42     15       1     79     15      0      1 0.00327   -2.16 0.00381
## 43      0       0     35     13      0      0 0.00154   -2.53 0.00246
## 44      0       0     50     16      0      1 0.00196   -2.45 0.00293
## 45     40       1     46      6      1      0 0.00620   -1.36 0.00289
## 46      0       0     78     16      0      0 0.00407   -2.50 0.00636
## 47      0       0     57     16      0      0 0.00220   -2.46 0.00332
## 48     15       1     42     17      0      0 0.00223   -2.01 0.00226
## 49     30       1     73      8      1      0 0.00456   -1.77 0.00357
## 50      0       0     22     15      0      1 0.00260   -2.43 0.00384
## 51      0       0     78     12      0      1 0.00319   -2.65 0.00560
## 52      0       0     72      9      0      0 0.00392   -2.76 0.00745
## 53      0       0     62     14      0      1 0.00176   -2.54 0.00285
## 54      0       1     91     14      0      1 0.00474   -2.87 0.00976
## 55      0       0     46     15      0      1 0.00150   -2.48 0.00230
## 56      0       0     54     17      0      1 0.00269   -2.41 0.00392
## 57    100       0     72      6      1      0 0.00698    1.46 0.00374
## 58      0       1     58     12      0      0 0.00134   -2.88 0.00277
## 59      0       0     65     11      0      1 0.00227   -2.67 0.00402
```

```
## 60   100     0   75    6     1     0 0.00723    1.46 0.00385
## 61     0     0   63   17     0     0 0.00320   -2.43 0.00474
## 62    15     1   66   16     0     1 0.00241   -2.09 0.00264
## 63     0     1   34   14     0     1 0.00140   -2.76 0.00266
## 64    15     0   62   17     0     0 0.00313   -1.78 0.00248
## 65    10     0   46   17     0     1 0.00250   -1.97 0.00242
## 66   100     0   33   17     1     0 0.00274    1.95 0.00261
## 67     0     0   77   16     0     1 0.00394   -2.50 0.00615
## 68     0     0   62   14     0     1 0.00176   -2.54 0.00285
## 69    15     0   24   12     0     0 0.00252   -1.90 0.00228
## 70     0     0   39   12     0     0 0.00155   -2.58 0.00258
## 71     0     1   66   17     0     0 0.00305   -2.71 0.00558
## 72    15     1   69   14     0     0 0.00193   -2.17 0.00229
## 73     0     0   32   16     1     0 0.00223   -2.41 0.00324
## 74   100     0   83    4     1     0 0.01098    1.37 0.00518
## 75    15     0   72   12     0     1 0.00255   -1.99 0.00252
## 76   100     0   82   11     1     0 0.00393    1.63 0.00263
## 77   100     0   80   12     1     0 0.00343    1.67 0.00241
## 78     0     1   33   13     0     0 0.00148   -2.80 0.00289
## 79     0     1   24   15     0     0 0.00227   -2.71 0.00415
## 80     0     1   45   12     0     1 0.00122   -2.86 0.00248
## 81     0     0   27   14     0     0 0.00202   -2.48 0.00310
## 82     0     0   77   16     0     1 0.00394   -2.50 0.00615
## 83    15     1   57   17     0     0 0.00248   -2.04 0.00257
## 84   100     1   91   12     1     0 0.00463    1.39 0.00224
## 85   100     0   85    3     1     0 0.01295    1.33 0.00576
## 86    15     1   24   16     0     0 0.00260   -2.02 0.00264
## 87     0     1   62   14     0     0 0.00144   -2.81 0.00284
## 88   100     0   78    9     1     0 0.00450    1.56 0.00276
## 89     0     0   23   11     0     1 0.00303   -2.59 0.00508
## 90     0     0   70   12     1     0 0.00236   -2.64 0.00411
```

Looking at the abnormal observations, we can see that most of these onbservations have extreme evaluation of biden thermometer. These observations predominantly have 0 evaluations of their biden warmth index. And most of these observations are republican. Then the outliers appear because we have omitted variables. Moving forward, I would add addtional variables in the model. For this case, I would add party affiliation to the model. It's intuitive since that with the same demographic background (in this case, same education, gender and age), they will differ significantly in their ideology.

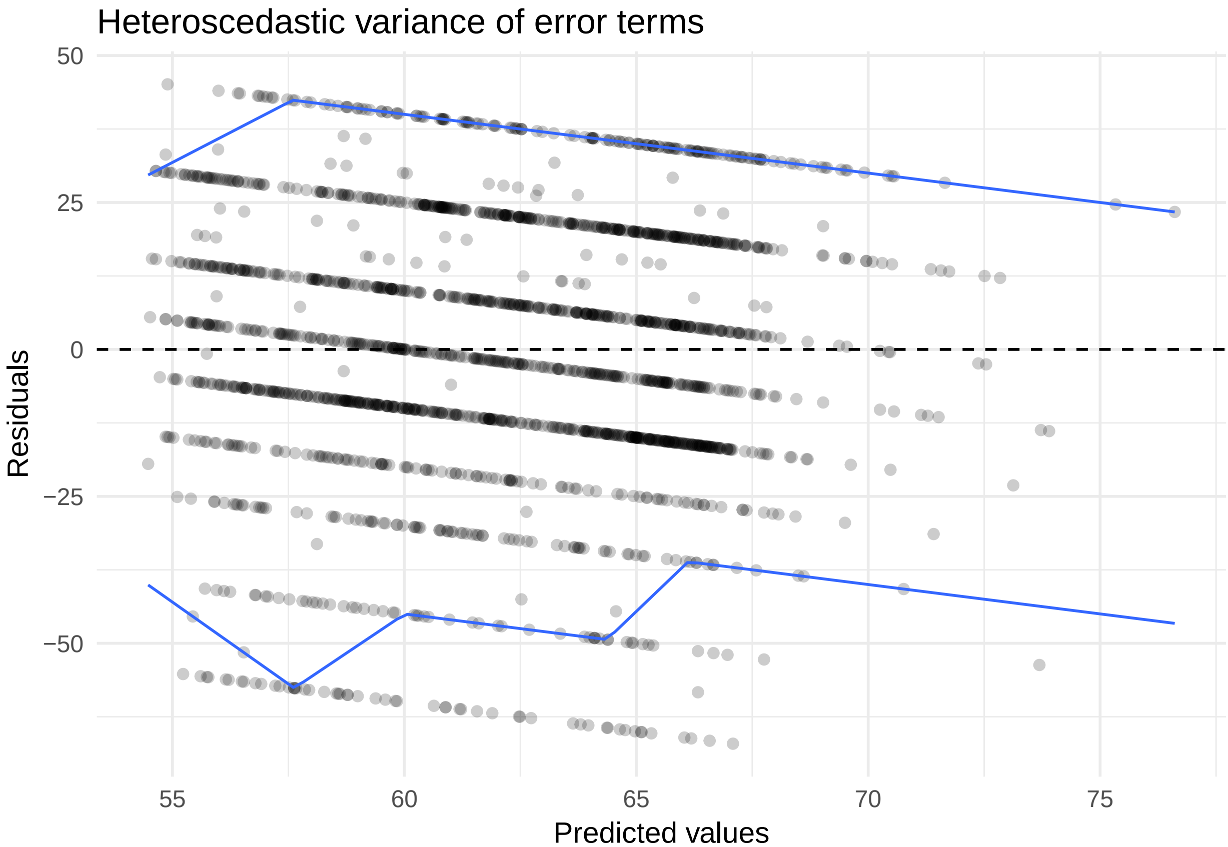
b)





According to the qqplot and the density plot of studendized residuals, we can see that the density plot is highly skewed. With the skewness of this plot, we can do a log transformation of the dependent variable.

c)



We can see that there is a problem of heterskcdesticity. For this problem, I would use weighted least square or generalized least square to control for the difference in variance.

d)

We can directly look at the VIF value for each variable to see if there is multicolinearity in the model.

```
##    age female   educ
##   1.01   1.00   1.01
```

All of the variables have value around 1 which is away from the threshold value 10. Then there is in general no multicolinearity problem in the model.

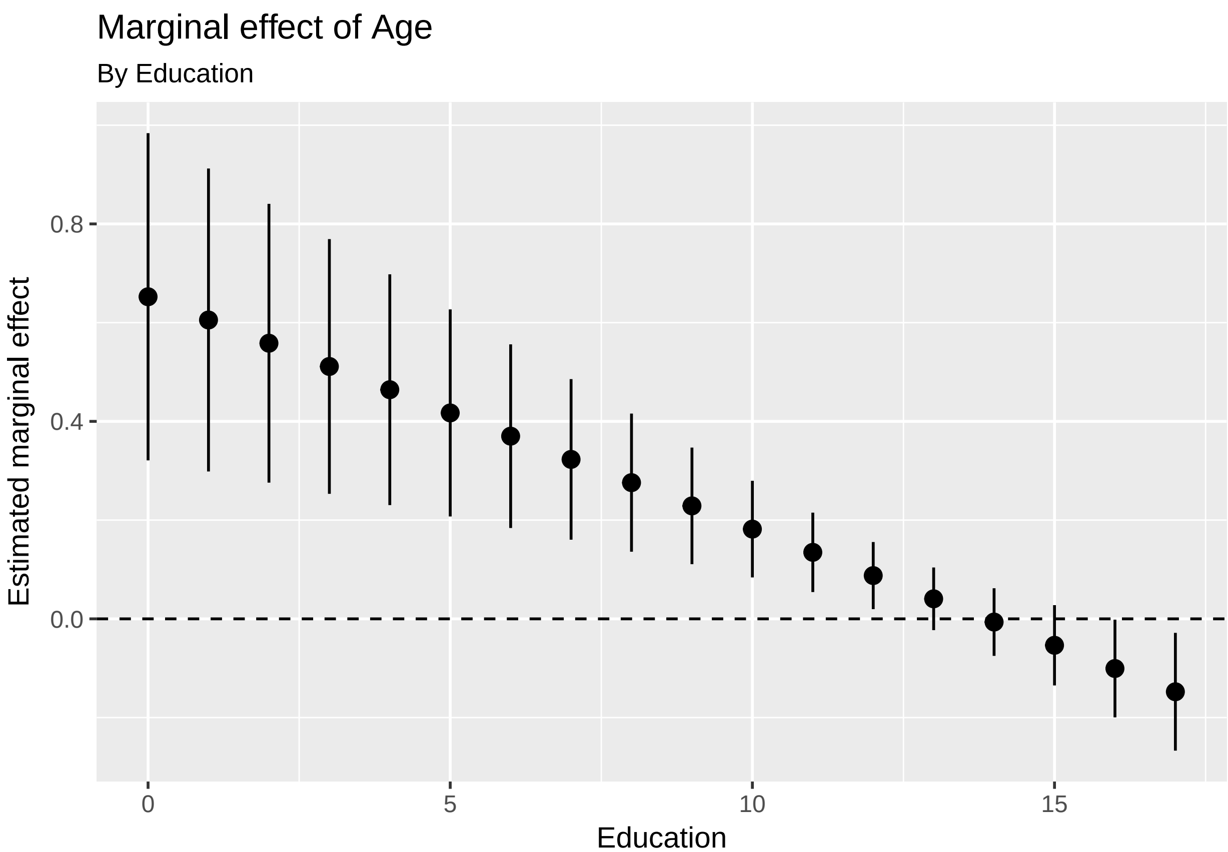
Problem 2

a)

```
##          term    estimate  std.error statistic      p.value
## 1 (Intercept) 36.29088878 9.49072252  3.823828 1.358657e-04
## 2         age  0.65245883 0.16909917  3.858439 1.181316e-04
## 3      female  6.14218010 1.09300304  5.619545 2.214101e-08
## 4        educ  1.58273997 0.70813807  2.235073 2.553470e-02
```

```
## 5    age:educ -0.04707047 0.01279497 -3.678826 2.411904e-04
```

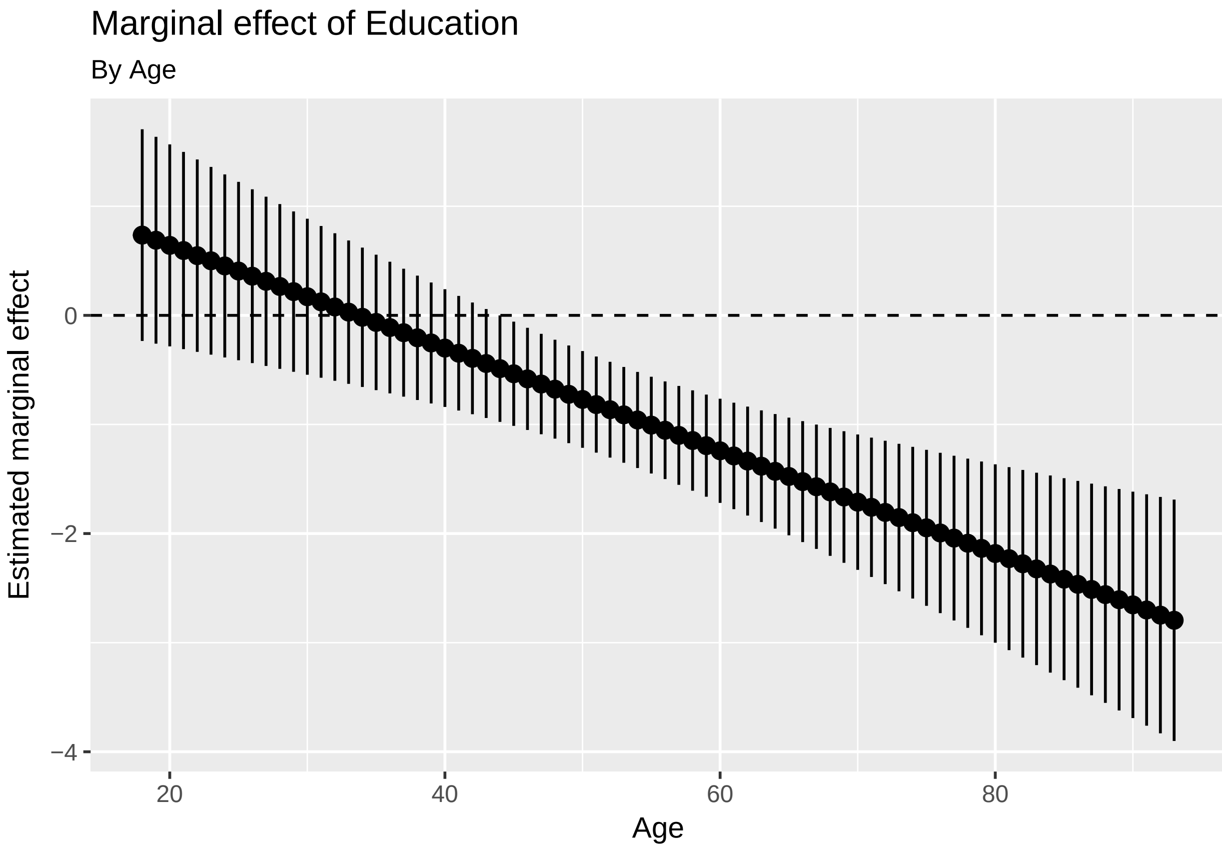
```
## Linear hypothesis test
##
## Hypothesis:
## age  + age:educ = 0
##
## Model 1: restricted model
## Model 2: biden ~ age + female + educ + educ * age
##
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1   1803 967831
## 2   1802 959867  1    7964.6 14.952 0.0001142 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



b)

```
## Linear hypothesis test
##
## Hypothesis:
## educ  + age:educ = 0
##
## Model 1: restricted model
## Model 2: biden ~ age + female + educ + educ * age
##
```

```
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1   1803 962460
## 2   1802 959867  1    2593.1 4.8681 0.02748 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

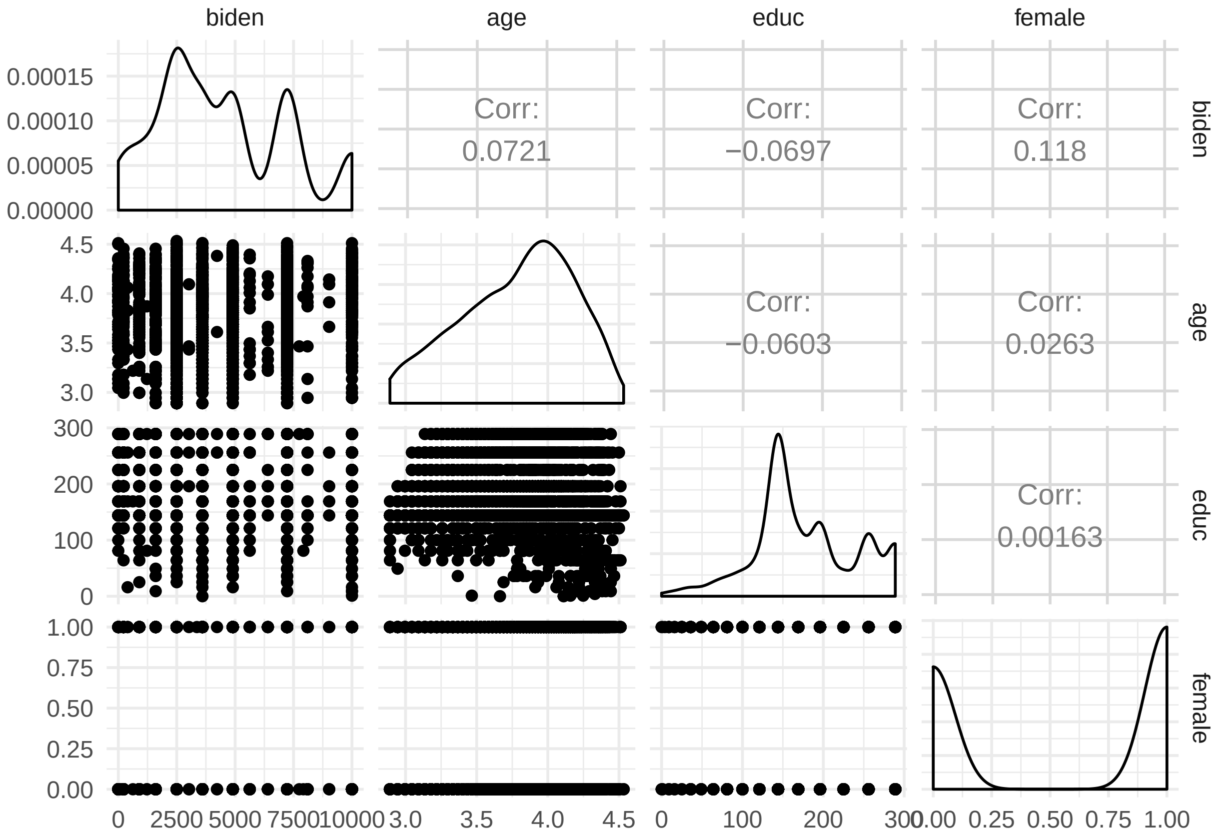


The above is the statistical inferene results. We can see that the model is statistically significant in 0.05 level. The effect of education is slightly less robust than the effect of age. This happens because the education effect

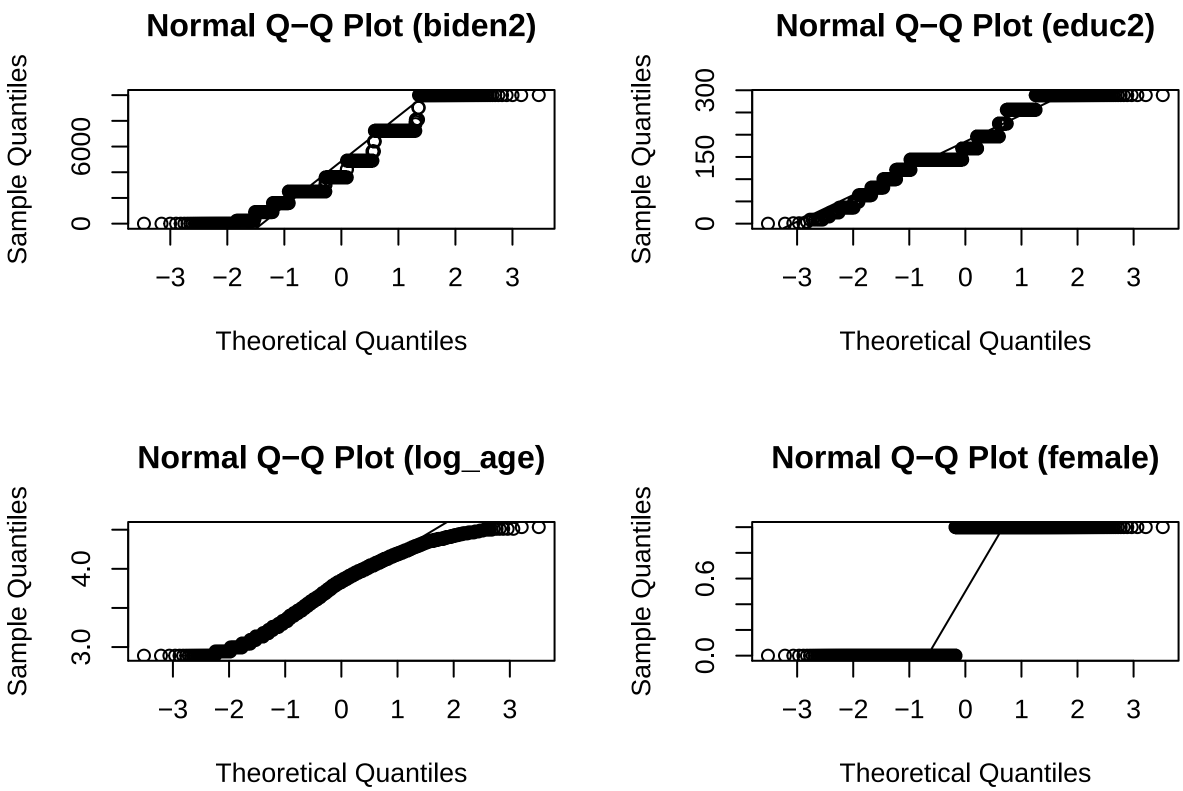
Problem 3

a)

We firstly plot the scatterplot matrix to visualize the distribution of each variable.



Now we will transform the raw data set to make each variable close to a normal distribution. Judging by the distribution of each variable. We perform log transformation on age; and we change education and the biden warmth variable to quadratic term to adjust for the skewness of the variables. As a result, we can see that education and age variable do conform to normal distribution more compared to the original variable.




```
## -- Imputation 1 --
##
##   1  2  3  4  5
##
## -- Imputation 2 --
##
##   1  2  3  4  5
##
## -- Imputation 3 --
##
##   1  2  3  4  5
##
## -- Imputation 4 --
##
##   1  2  3  4
##
## -- Imputation 5 --
##
##   1  2  3  4  5
```

```
##          term    estimate  std.error statistic      p.value
## 1 (Intercept) 68.62101396 3.59600465 19.082571 4.337464e-74
## 2         age  0.04187919 0.03248579  1.289154 1.975099e-01
## 3      female  6.19606946 1.09669702  5.649755 1.863612e-08
## 4        educ -0.88871263 0.22469183 -3.955251 7.941295e-05
```

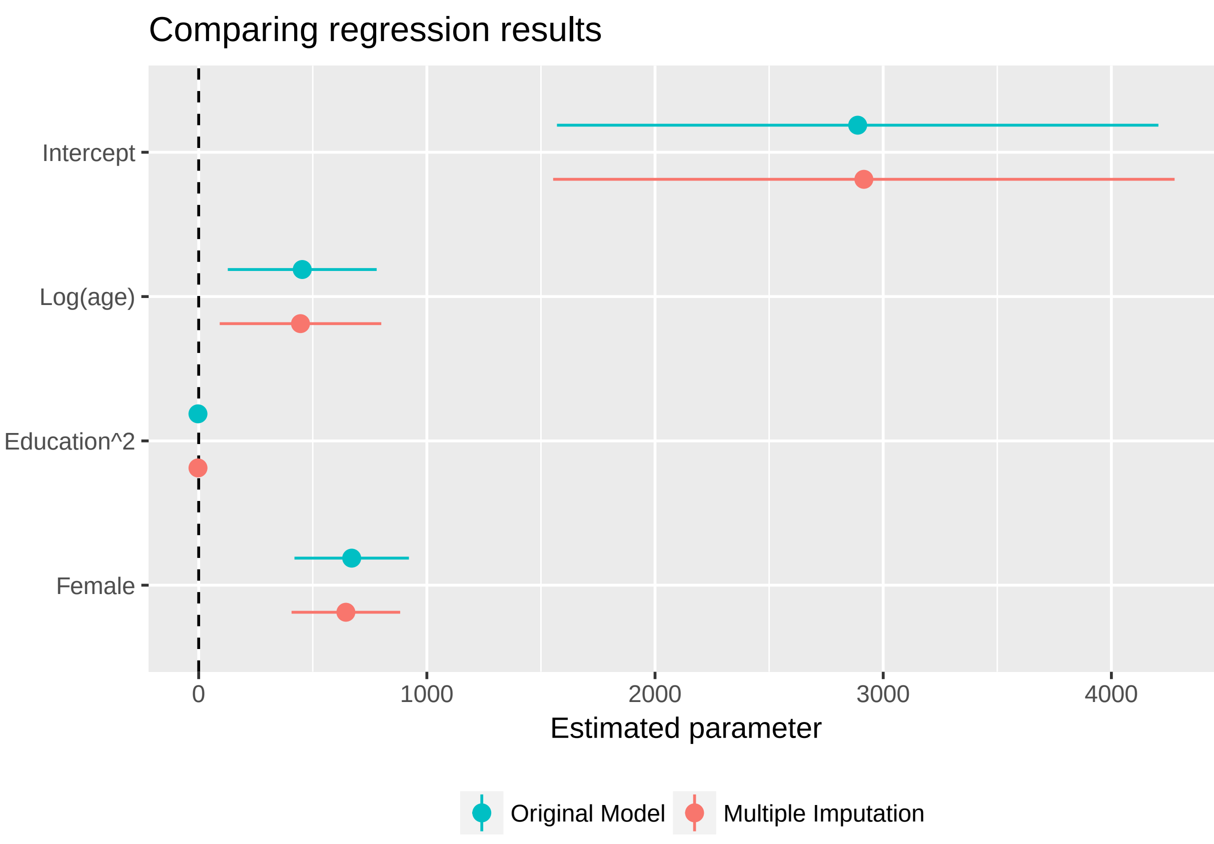
Now we display the estimates of the imputation.

```
## # A tibble: 20 × 6
##       id        term    estimate   std.error statistic      p.value
```

```
##     <chr>       <chr>        <dbl>       <dbl>     <dbl>        <dbl>
## 1   imp1 (Intercept) 3427.378526 575.6821611  5.953595 3.022100e-09
## 2   imp1     log_age  329.600819 142.7958559  2.308196 2.107583e-02
## 3   imp1       educ2   -3.627558   0.8744211 -4.148526 3.466804e-05
## 4   imp1      female  673.087413 112.8201654  5.966020 2.804056e-09
## 5   imp2 (Intercept) 2781.591657 579.2884963  4.801738 1.673218e-06
## 6   imp2     log_age  464.823209 143.4941877  3.239317 1.215093e-03
## 7   imp2       educ2   -2.754921   0.8792863 -3.133133 1.751116e-03
## 8   imp2      female  636.501894 113.3390676  5.615909 2.189095e-08
## 9   imp3 (Intercept) 2691.540115 587.0992031  4.584472 4.793163e-06
## 10  imp3     log_age  514.178409 145.3789467  3.536815 4.129310e-04
## 11  imp3       educ2   -3.002686   0.8906295 -3.371420 7.600803e-04
## 12  imp3      female  608.459994 114.8459010  5.298056 1.280891e-07
## 13  imp4 (Intercept) 3104.434530 580.3078210  5.349634 9.678209e-08
## 14  imp4     log_age  359.321993 143.6676707  2.501064 1.245056e-02
## 15  imp4       educ2   -2.610509   0.8838195 -2.953668 3.171806e-03
## 16  imp4      female  696.304236 113.7977462  6.118788 1.103658e-09
## 17  imp5 (Intercept) 2571.259004 582.0400410  4.417667 1.043586e-05
## 18  imp5     log_age  562.966627 144.1904420  3.904327 9.718540e-05
## 19  imp5       educ2   -3.295891   0.8855372 -3.721912 2.024250e-04
## 20  imp5      female  610.881345 114.3047529  5.344321 9.962813e-08
```

The following is the comparison of the full model and the imputed model.

```
##          term    estimate  std.error estimate.mi std.error.mi
## 1 (Intercept) 2888.096643 672.425266 2915.240766  694.7889880
## 2     log_age  453.836614 166.488616  446.178212  180.6465010
## 3       educ2   -3.148178   1.017445   -3.058313    0.9910117
## 4      female  670.576237 127.946727  645.046977  121.4643583
```



We can see that the difference between the two models are not large. The imputation model has slightly larger standard error for estimators. The model indicates that age and gender are more important determinant of the biden warmth index.