Analysis of Income by College Major

Saif Ul Mehdi

02/07/2020

Dependencies

Loading the Dependencies

```
library(collegeIncome)
data(college)
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
```

Exploratory Data Analysis

Finding what does the data have

```
head(college)
```

```
##   rank major_code                                     major major_category
## 1    1       2419                     Petroleum Engineering    Engineering
## 2    2       2416            Mining And Mineral Engineering    Engineering
## 3    3       2415                 Metallurgical Engineering    Engineering
## 4    4       2417 Naval Architecture And Marine Engineering    Engineering
## 5    5       2405                      Chemical Engineering    Engineering
## 6    6       2418                       Nuclear Engineering    Engineering
##   total sample_size perc_women p25th median  p75th   perc_men perc_employed
## 1  2339          36  0.9109326 25000  40000  50000 0.08906743     0.9115044
## 2   756           7  0.5154064 26000  37000  40000 0.48459355     0.7980501
## 3   856           3  0.5942076 26700  45000  60000 0.40579235     0.7871943
## 4  1258          16  0.6521298 26000  35000  45000 0.34787018     0.8465608
## 5 32260         289  0.4179248 31500  62000 109000 0.58207520     0.8515625
## 6  2573          17  0.4305368 23000  44700  50000 0.56946324     0.8474507
```

```
##   perc_employed_fulltime perc_employed_parttime
## 1              0.9206524              0.1774785
## 2              0.7110092              0.3623853
## 3              0.8833498              0.3387257
## 4              0.9366337              0.1673267
## 5              0.8086363              0.4020061
## 6              0.8756262              0.2040405
##   perc_employed_fulltime_yearround perc_unemployed perc_college_jobs
## 1                        0.7704431      0.08849558         0.6702970
## 2                        0.7093101      0.20194986         0.3867764
## 3                        0.7738366      0.21280567         0.7289116
## 4                        0.6527853      0.15343915         0.2460902
## 5                        0.6852821      0.14843750         0.5867515
## 6                        0.6567727      0.15254929         0.4624782
##   perc_non_college_jobs perc_low_wage_jobs
## 1             0.1821782         0.05544554
## 2             0.5158761         0.21560172
## 3             0.1759983         0.03014828
## 4             0.4107636         0.04323827
## 5             0.3860437         0.11801062
## 6             0.4057592         0.23472949
```

```r
tail(college)
```

```
##     rank major_code                    major            major_category total
## 168  168       3302 Composition And Rhetoric Humanities & Liberal Arts 18953
## 169  169       3609                  Zoology    Biology & Life Science  8409
## 170  170       5201   Educational Psychology  Psychology & Social Work  2854
## 171  171       5202      Clinical Psychology  Psychology & Social Work  2838
## 172  172       5203    Counseling Psychology  Psychology & Social Work  4626
## 173  173       3501          Library Science                 Education  1098
##     sample_size perc_women p25th median  p75th  perc_men perc_employed
## 168         151  0.8459344 30000  42000  65000 0.1540656     0.7636511
## 169          47  0.7643203 50000  65000 102000 0.2356797     0.6757741
## 170           7  0.8644561 33000  46000  58000 0.1355439     0.7932137
## 171          13  0.8128766 22000  29000  38000 0.1871234     0.8017061
## 172          21  0.5847764 39000  48000  58000 0.4152236     0.7403101
## 173           2  0.3212961 22500  38400  45000 0.6787039     0.8194622
##     perc_employed_fulltime perc_employed_parttime
## 168              1.0041209              0.1016484
## 169              0.8792842              0.1889597
## 170              0.9613045              0.1179815
## 171              0.8414807              0.2807614
## 172              0.8203650              0.2846461
## 173              0.7470044              0.3622428
##     perc_employed_fulltime_yearround perc_unemployed perc_college_jobs
## 168                        0.7687849       0.2363489         0.6798525
## 169                        0.6058012       0.3242259         0.3260464
## 170                        0.7406321       0.2067863         0.3928227
## 171                        0.7271024       0.1982939         0.2131006
## 172                        0.7809422       0.2596899         0.3483973
## 173                        0.6835719       0.1805378         0.7803185
##     perc_non_college_jobs perc_low_wage_jobs
## 168             0.2782434         0.08716058
```

```
## 169             0.5193282         0.05145295
## 170             0.4748271         0.13746574
## 171             0.5087367         0.15915810
## 172             0.5483503         0.19906500
## 173             0.1245406         0.02858310
```

```r
summary(college)
```

```
##       rank        major_code        major           major_category    
##  Min.   :  1   Min.   :1100   Length:173         Length:173        
##  1st Qu.: 44   1st Qu.:2403   Class :character   Class :character  
##  Median : 87   Median :3608   Mode  :character   Mode  :character  
##  Mean   : 87   Mean   :3880                                        
##  3rd Qu.:130   3rd Qu.:5503                                        
##  Max.   :173   Max.   :6403                                        
##                                                                    
##      total         sample_size       perc_women          p25th      
##  Min.   :   124   Min.   :   2.0   Min.   :0.0000   Min.   :18500  
##  1st Qu.:  4361   1st Qu.:  39.0   1st Qu.:0.3397   1st Qu.:24000  
##  Median : 15058   Median : 130.0   Median :0.5357   Median :27000  
##  Mean   : 39168   Mean   : 356.1   Mean   :0.5226   Mean   :29501  
##  3rd Qu.: 38844   3rd Qu.: 338.0   3rd Qu.:0.7020   3rd Qu.:33000  
##  Max.   :393735   Max.   :4212.0   Max.   :0.9690   Max.   :95000  
##                                                                   
##      median           p75th           perc_men        perc_employed   
##  Min.   : 22000   Min.   : 22000   Min.   :0.03105   Min.   :0.0000  
##  1st Qu.: 33000   1st Qu.: 42000   1st Qu.:0.29798   1st Qu.:0.7477  
##  Median : 36000   Median : 47000   Median :0.46429   Median :0.8028  
##  Mean   : 40151   Mean   : 51494   Mean   :0.47745   Mean   :0.7886  
##  3rd Qu.: 45000   3rd Qu.: 60000   3rd Qu.:0.66033   3rd Qu.:0.8410  
##  Max.   :110000   Max.   :125000   Max.   :1.00000   Max.   :0.9562  
##                                                                      
##  perc_employed_fulltime perc_employed_parttime perc_employed_fulltime_yearround
##  Min.   :0.5743         Min.   :0.0000         Min.   :0.5857                  
##  1st Qu.:0.7741         1st Qu.:0.2090         1st Qu.:0.7009                  
##  Median :0.8319         Median :0.2862         Median :0.7484                  
##  Mean   :   Inf         Mean   :0.2874         Mean   :0.7476                  
##  3rd Qu.:0.8974         3rd Qu.:0.3623         3rd Qu.:0.7896                  
##  Max.   :   Inf         Max.   :0.5518         Max.   :1.0000                  
##                         NA's   :1                                              
##  perc_unemployed   perc_college_jobs perc_non_college_jobs perc_low_wage_jobs
##  Min.   :0.04383   Min.   :0.0633    Min.   :0.08278       Min.   :0.00000   
##  1st Qu.:0.15899   1st Qu.:0.2974    1st Qu.:0.27995       1st Qu.:0.06957   
##  Median :0.19723   Median :0.4160    Median :0.42020       Median :0.10857   
##  Mean   :0.21140   Mean   :0.4478    Mean   :0.41498       Mean   :0.11481   
##  3rd Qu.:0.25229   3rd Qu.:0.6170    3rd Qu.:0.52756       3rd Qu.:0.15353   
##  Max.   :1.00000   Max.   :0.8383    Max.   :0.85364       Max.   :0.36566   
##                    NA's   :1         NA's   :1             NA's   :1
```

Different categories of college majors

```r
table(college$major_category)
```

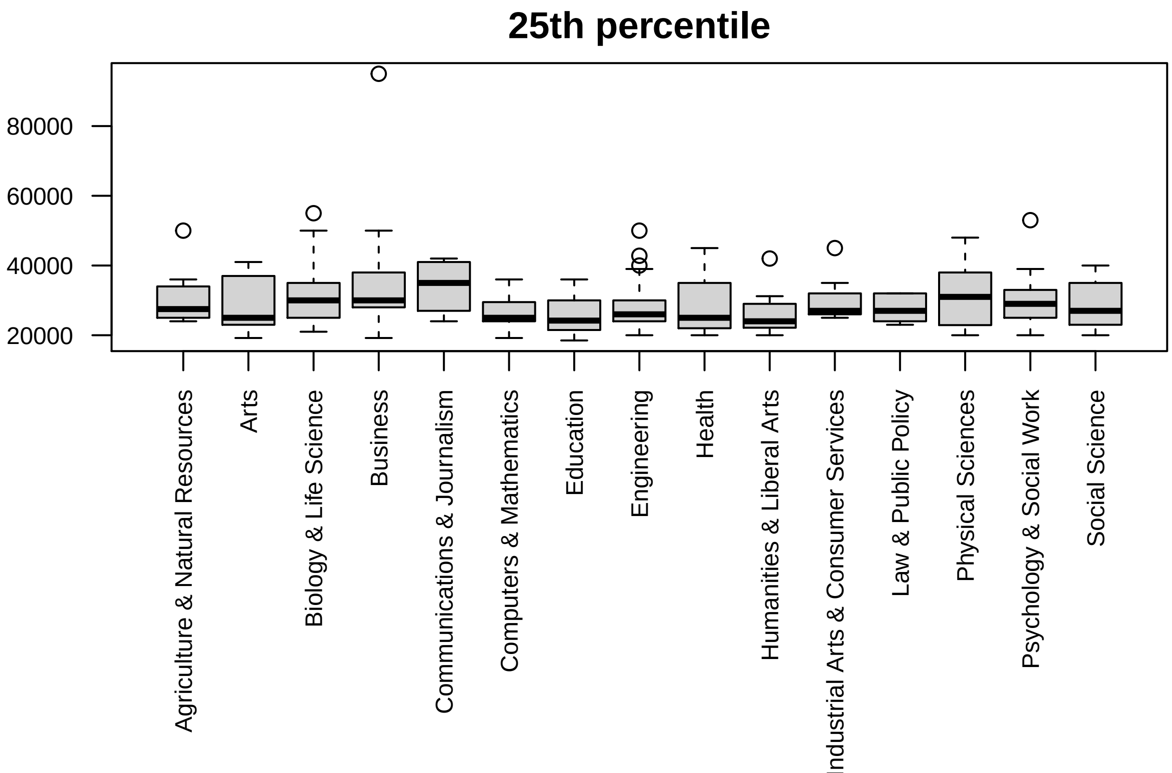
```
##
##     Agriculture & Natural Resources                                Arts
##                                  10                                   8
##              Biology & Life Science                            Business
##                                  14                                  13
##         Communications & Journalism            Computers & Mathematics
##                                   4                                  11
##                           Education                         Engineering
##                                  16                                  29
##                              Health           Humanities & Liberal Arts
##                                  12                                  15
## Industrial Arts & Consumer Services                   Interdisciplinary
##                                   7                                   1
##                 Law & Public Policy                   Physical Sciences
##                                   5                                  10
##           Psychology & Social Work                      Social Science
##                                   9                                   9
```

Interdisciplinary college major will not be helpful in estimation so, we will remove it from the dataset

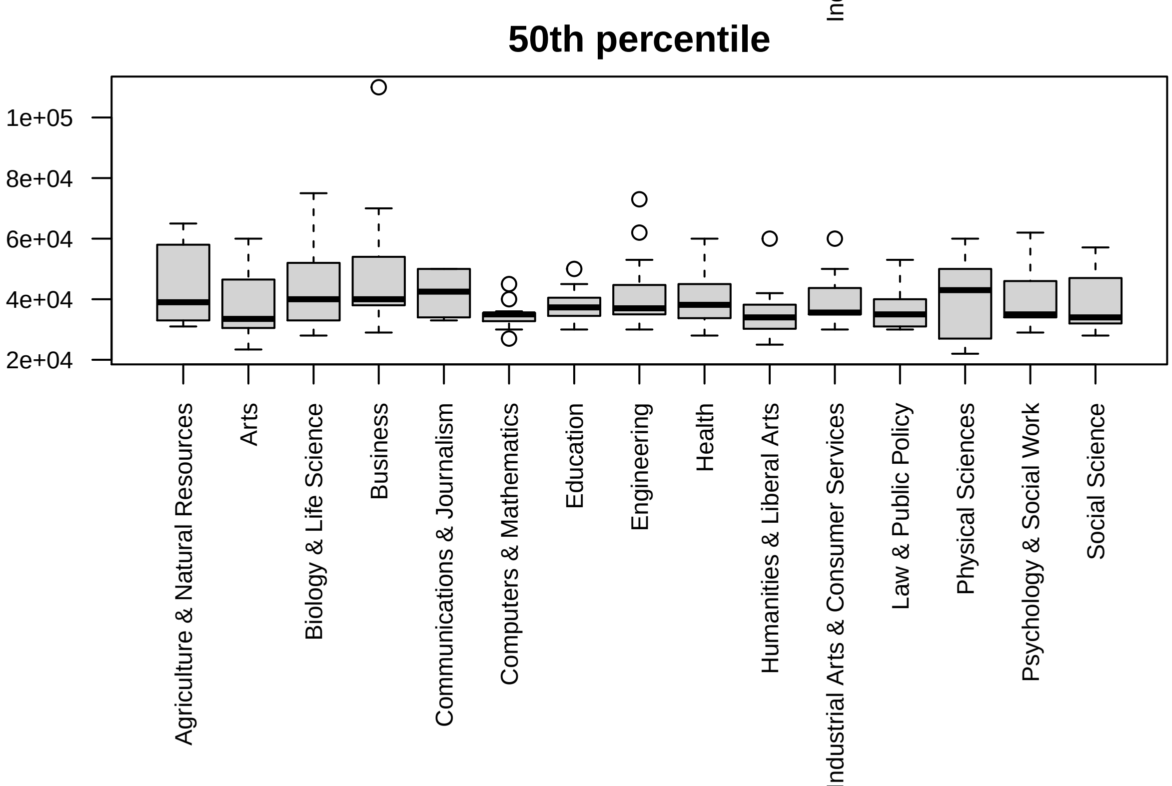
```r
college<-college%>%filter(major_category!='Interdisciplinary')
table(college$major_category)
```

```
##
##     Agriculture & Natural Resources                                Arts
##                                  10                                   8
##              Biology & Life Science                            Business
##                                  14                                  13
##         Communications & Journalism            Computers & Mathematics
##                                   4                                  11
##                           Education                         Engineering
##                                  16                                  29
##                              Health           Humanities & Liberal Arts
##                                  12                                  15
## Industrial Arts & Consumer Services                 Law & Public Policy
##                                   7                                   5
##                   Physical Sciences           Psychology & Social Work
##                                  10                                   9
##                      Social Science
##                                   9
```

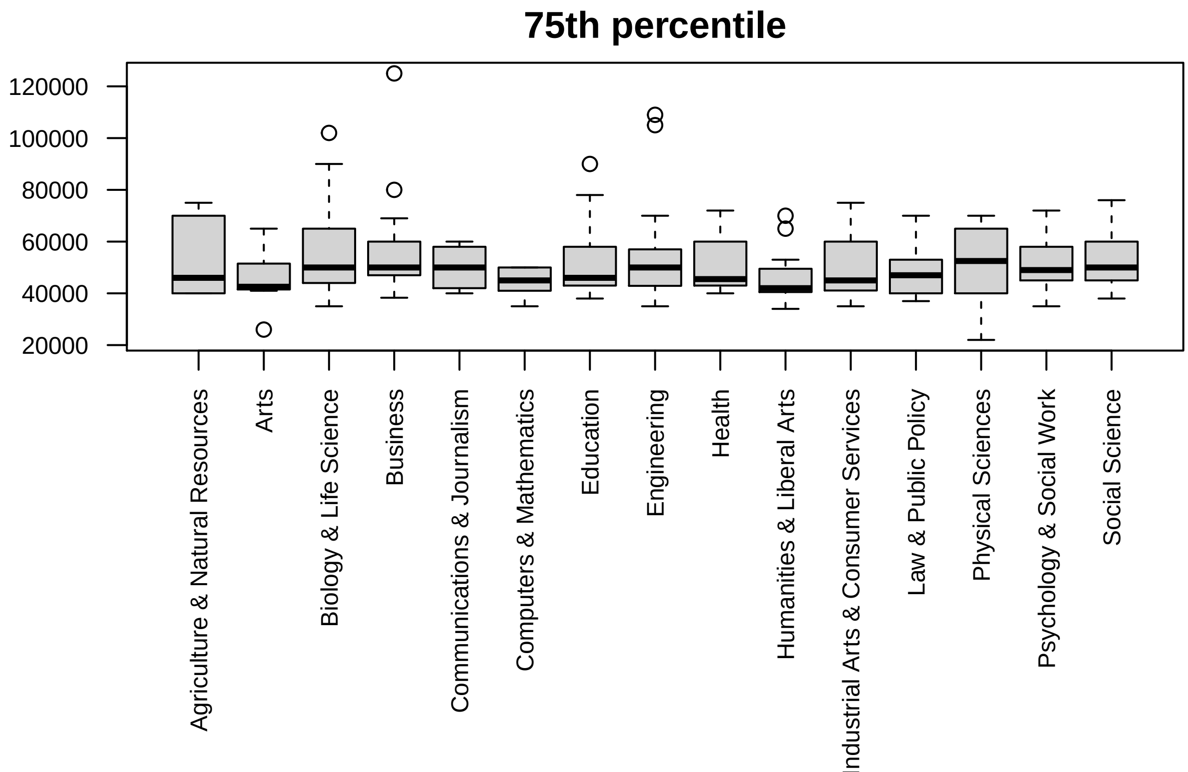
Box Plot



25th percentile
80000
60000
40000
20000
Agriculture & Natural Resources
Arts
Biology & Life Science
Business
Communications & Journalism
Computers & Mathematics
Education
Engineering
Health
Humanities & Liberal Arts
Industrial Arts & Consumer Services
Law & Public Policy
Physical Sciences
Psychology & Social Work
Social Science



50th percentile
1e+05
8e+04
6e+04
4e+04
2e+04
Agriculture & Natural Resources
Arts
Biology & Life Science
Business
Communications & Journalism
Computers & Mathematics
Education
Engineering
Health
Humanities & Liberal Arts
Industrial Arts & Consumer Services
Law & Public Policy
Physical Sciences
Psychology & Social Work
Social Science



There seems to be no considerable variation across categories for either of these three measures, so we will use median as our outcome measure.

Linear Regression

Since, the dataset consists of different types of jobs like low paying jobs, jobs which require college degree we would include it in the linear model. We will also include the gender effects also. We will fit the model and draw inferences from it

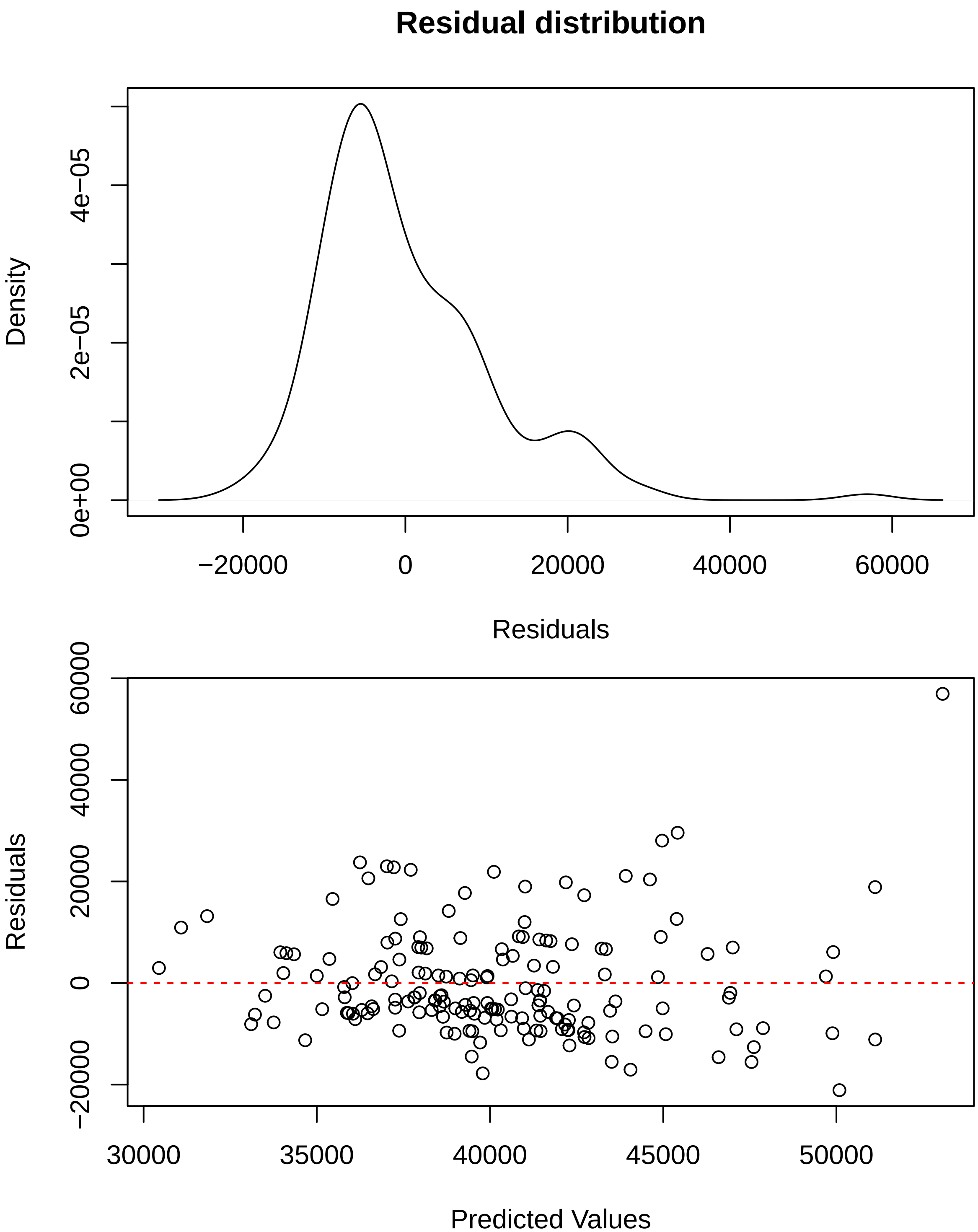
```
fit<-lm(median~major_category+perc_women+perc_college_jobs+perc_low_wage_jobs,college)
summary(fit)
```

```
##
## Call:
## lm(formula = median ~ major_category + perc_women + perc_college_jobs +
##     perc_low_wage_jobs, data = college)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -21088  -6917  -3211   5965  56932
##
## Coefficients:
##                                                    Estimate Std. Error t value
## (Intercept)                                         49491.3     5647.2   8.764
## major_categoryArts                                  -5429.2     5435.1  -0.999
## major_categoryBiology & Life Science                  707.2     4772.0   0.148
## major_categoryBusiness                               5540.4     4846.5   1.143
## major_categoryCommunications & Journalism           -2776.6     6761.1  -0.411
## major_categoryComputers & Mathematics               -9607.2     5129.6  -1.873
```

```
## major_categoryEducation                           -5140.6     4591.3  -1.120
## major_categoryEngineering                         -2931.4     4185.3  -0.700
## major_categoryHealth                              -3700.9     4880.2  -0.758
## major_categoryHumanities & Liberal Arts           -9022.7     4711.1  -1.915
## major_categoryIndustrial Arts & Consumer Services -2731.0     5604.6  -0.487
## major_categoryLaw & Public Policy                 -5612.6     6374.9  -0.880
## major_categoryPhysical Sciences                   -3268.4     5109.4  -0.640
## major_categoryPsychology & Social Work            -5102.7     5340.0  -0.956
## major_categorySocial Science                      -3059.1     5287.2  -0.579
## perc_women                                        -5011.5     3958.1  -1.266
## perc_college_jobs                                 -7841.4     5283.5  -1.484
## perc_low_wage_jobs                                 2108.6    16018.8   0.132
##                                                   Pr(>|t|)
## (Intercept)                                       3.36e-15 ***
## major_categoryArts                                  0.3194
## major_categoryBiology & Life Science                0.8824
## major_categoryBusiness                              0.2548
## major_categoryCommunications & Journalism           0.6819
## major_categoryComputers & Mathematics               0.0630 .
## major_categoryEducation                             0.2646
## major_categoryEngineering                           0.4847
## major_categoryHealth                                0.4494
## major_categoryHumanities & Liberal Arts             0.0573 .
## major_categoryIndustrial Arts & Consumer Services   0.6268
## major_categoryLaw & Public Policy                   0.3800
## major_categoryPhysical Sciences                     0.5233
## major_categoryPsychology & Social Work              0.3408
## major_categorySocial Science                        0.5637
## perc_women                                          0.2074
## perc_college_jobs                                   0.1398
## perc_low_wage_jobs                                  0.8954
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11330 on 153 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.1236, Adjusted R-squared:  0.02627
## F-statistic:  1.27 on 17 and 153 DF,  p-value: 0.2191
```

Looking at the summary of the fitted models, it is obvious that the categories of majors don't have effect on the median income as the F-statistic. We would gain further inference by drawing diagnostic plots

Diagnostic Plots



We can see from the diagnostic plots that the assumption of residuals being normal is valid and heteroskedasticity is not an issue. There could be a possible outlier. Overall, there seems to be no effect of college major category on median income