College Major vs Income

Himank Jain

27/12/2019

College Major vs Income:

We try to answer below questions:

Are incomes of individuals from different majors in college differ significantly from each other or not? If so , Which Major pays higher and which pays Lower?

```
#devtools::install_github("jhudsl/collegeIncome")
library(collegeIncome)
#devtools::install_github("jhudsl/matahari")
library(matahari)
```

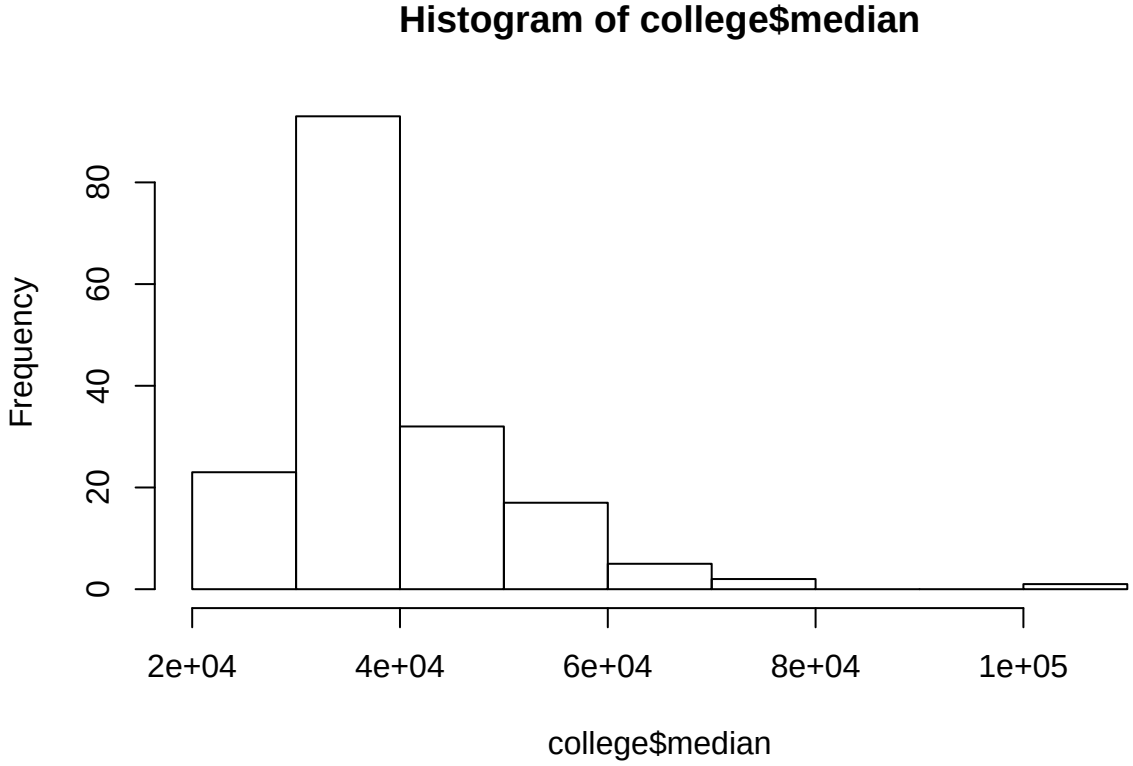
EDA:

How does the data look like?

```
##   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
```

```
## 'data.frame':    173 obs. of  19 variables:
##  $ rank                           : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ major_code                     : int  2419 2416 2415 2417 2405 2418 6202 5001 2414 2408 ...
##  $ major                          : chr  "Petroleum Engineering" "Mining And Mineral Engineering" "
##  $ major_category                 : chr  "Engineering" "Engineering" "Engineering" "Engineering" ..
##  $ total                          : int  2339 756 856 1258 32260 2573 3777 1792 91227 81527 ...
##  $ sample_size                    : int  36 7 3 16 289 17 51 10 1029 631 ...
##  $ perc_women                     : num  0.911 0.515 0.594 0.652 0.418 ...
##  $ p25th                          : num  25000 26000 26700 26000 31500 23000 32500 37900 29200 2300
##  $ median                         : num  40000 37000 45000 35000 62000 44700 45000 57000 36000 3220
##  $ p75th                          : num  50000 40000 60000 45000 109000 50000 58000 67000 46000 471
##  $ perc_men                       : num  0.0891 0.4846 0.4058 0.3479 0.5821 ...
##  $ perc_employed                  : num  0.912 0.798 0.787 0.847 0.852 ...
##  $ perc_employed_fulltime         : num  0.921 0.711 0.883 0.937 0.809 ...
##  $ perc_employed_parttime         : num  0.177 0.362 0.339 0.167 0.402 ...
##  $ perc_employed_fulltime_yearround: num  0.77 0.709 0.774 0.653 0.685 ...
##  $ perc_unemployed                : num  0.0885 0.2019 0.2128 0.1534 0.1484 ...
##  $ perc_college_jobs              : num  0.67 0.387 0.729 0.246 0.587 ...
##  $ perc_non_college_jobs          : num  0.182 0.516 0.176 0.411 0.386 ...
##  $ perc_low_wage_jobs             : num  0.0554 0.2156 0.0301 0.0432 0.118 ...
```



Missing Values? and Aggregating mean of income with respect to college majors

```r
any(is.na(college$major_category))
```

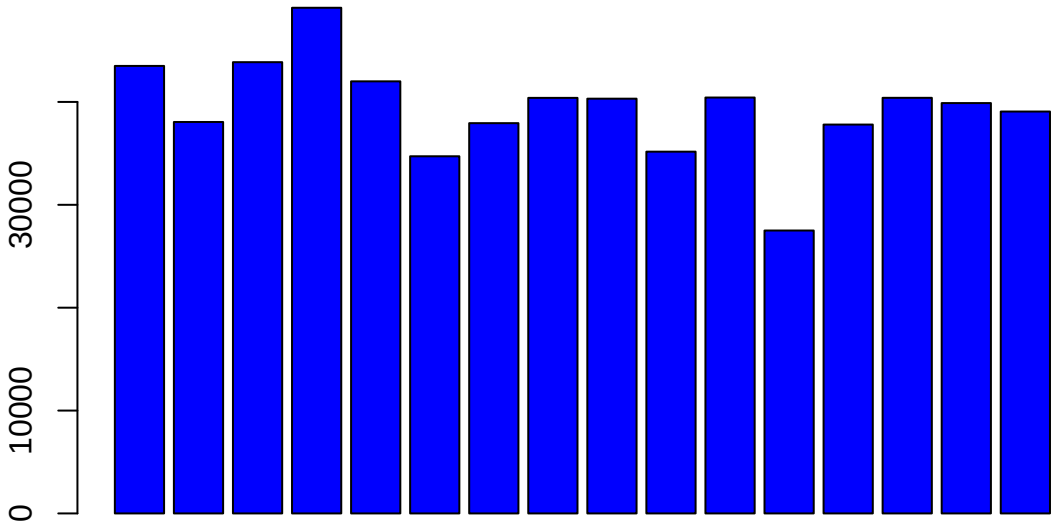
```
## [1] FALSE
```

```r
factor(college$major_category)
```

```
##   [1] Engineering                          Engineering
##   [3] Engineering                          Engineering
##   [5] Engineering                          Engineering
##   [7] Business                             Physical Sciences
##   [9] Engineering                          Engineering
##  [11] Engineering                          Engineering
##  [13] Engineering                          Engineering
##  [15] Engineering                          Engineering
##  [17] Engineering                          Engineering
##  [19] Engineering                          Law & Public Policy
##  [21] Computers & Mathematics              Agriculture & Natural Resources
##  [23] Engineering                          Engineering
##  [25] Business                             Engineering
##  [27] Industrial Arts & Consumer Services  Business
##  [29] Engineering                          Law & Public Policy
##  [31] Engineering                          Engineering
##  [33] Arts                                 Engineering
##  [35] Health                               Business
##  [37] Social Science                       Business
##  [39] Engineering                          Physical Sciences
##  [41] Business                             Computers & Mathematics
##  [43] Computers & Mathematics              Physical Sciences
##  [45] Health                               Computers & Mathematics
##  [47] Computers & Mathematics              Computers & Mathematics
##  [49] Biology & Life Science               Physical Sciences
##  [51] Engineering                          Health
##  [53] Computers & Mathematics              Computers & Mathematics
##  [55] Biology & Life Science               Education
##  [57] Social Science                       Business
##  [59] Engineering                          Business
##  [61] Health                               Biology & Life Science
##  [63] Business                             Agriculture & Natural Resources
##  [65] Agriculture & Natural Resources      Engineering
##  [67] Engineering                          Biology & Life Science
##  [69] Social Science                       Humanities & Liberal Arts
##  [71] Psychology & Social Work             Agriculture & Natural Resources
##  [73] Physical Sciences                    Industrial Arts & Consumer Services
##  [75] Physical Sciences                    Industrial Arts & Consumer Services
##  [77] Business                             Business
##  [79] Social Science                       Social Science
##  [81] Biology & Life Science               Computers & Mathematics
##  [83] Biology & Life Science               Biology & Life Science
##  [85] Computers & Mathematics              Physical Sciences
##  [87] Business                             Law & Public Policy
##  [89] Health                               Law & Public Policy
##  [91] Physical Sciences                    Psychology & Social Work
```

```
##  [93] Biology & Life Science              Communications & Journalism
##  [95] Law & Public Policy                 Arts
##  [97] Communications & Journalism         Physical Sciences
##  [99] Communications & Journalism         Humanities & Liberal Arts
## [101] Education                           Biology & Life Science
## [103] Social Science                      Health
## [105] Health                              Computers & Mathematics
## [107] Industrial Arts & Consumer Services Agriculture & Natural Resources
## [109] Biology & Life Science              Interdisciplinary
## [111] Physical Sciences                   Agriculture & Natural Resources
## [113] Agriculture & Natural Resources     Education
## [115] Humanities & Liberal Arts           Humanities & Liberal Arts
## [117] Humanities & Liberal Arts           Education
## [119] Health                              Education
## [121] Education                           Health
## [123] Biology & Life Science              Biology & Life Science
## [125] Social Science                      Communications & Journalism
## [127] Health                              Business
## [129] Education                           Humanities & Liberal Arts
## [131] Education                           Social Science
## [133] Biology & Life Science              Education
## [135] Health                              Humanities & Liberal Arts
## [137] Education                           Humanities & Liberal Arts
## [139] Education                           Industrial Arts & Consumer Services
## [141] Humanities & Liberal Arts           Arts
## [143] Social Science                      Agriculture & Natural Resources
## [145] Education                           Psychology & Social Work
## [147] Arts                                Education
## [149] Humanities & Liberal Arts           Arts
## [151] Industrial Arts & Consumer Services Psychology & Social Work
## [153] Agriculture & Natural Resources     Arts
## [155] Education                           Psychology & Social Work
## [157] Psychology & Social Work            Humanities & Liberal Arts
## [159] Humanities & Liberal Arts           Arts
## [161] Industrial Arts & Consumer Services Agriculture & Natural Resources
## [163] Humanities & Liberal Arts           Health
## [165] Education                           Humanities & Liberal Arts
## [167] Arts                                Humanities & Liberal Arts
## [169] Biology & Life Science              Psychology & Social Work
## [171] Psychology & Social Work            Psychology & Social Work
## [173] Education
## 16 Levels: Agriculture & Natural Resources Arts ... Social Science
```

```r
df<-aggregate(college$median~college$major_category,FUN = mean)
names(df)<-c("major","medsal")
par(mfrow=c(1,1))
barplot(df$medsal,col='blue')
```



The Linear model:

```
fit<-lm(median~major_category,data=college)
summary(fit)$coef
```

```
##                                                  Estimate Std. Error
## (Intercept)                                   43500.0000   3590.819
## major_categoryArts                            -5450.0000   5386.228
## major_categoryBiology & Life Science            364.2857   4701.486
## major_categoryBusiness                         5653.8462   4776.236
## major_categoryCommunications & Journalism     -1500.0000   6717.807
## major_categoryComputers & Mathematics         -8781.8182   4961.429
## major_categoryEducation                       -5562.5000   4577.414
## major_categoryEngineering                     -3106.8966   4164.154
## major_categoryHealth                          -3183.3333   4861.992
## major_categoryHumanities & Liberal Arts       -8333.3333   4635.727
## major_categoryIndustrial Arts & Consumer Services  -3071.4286   5595.887
## major_categoryInterdisciplinary              -16000.0000  11909.399
## major_categoryLaw & Public Policy             -5700.0000   6219.481
## major_categoryPhysical Sciences               -3100.0000   5078.185
## major_categoryPsychology & Social Work        -3611.1111   5217.339
## major_categorySocial Science                  -4433.3333   5217.339
##                                                   t value     Pr(>|t|)
## (Intercept)                                   12.11422804 2.873928e-24
## major_categoryArts                            -1.01183974 3.131715e-01
## major_categoryBiology & Life Science           0.07748311 9.383379e-01
## major_categoryBusiness                         1.18374520 2.383031e-01
## major_categoryCommunications & Journalism     -0.22328715 8.236023e-01
## major_categoryComputers & Mathematics         -1.77001776 7.866520e-02
## major_categoryEducation                       -1.21520579 2.261119e-01
## major_categoryEngineering                     -0.74610504 4.567197e-01
## major_categoryHealth                          -0.65473851 5.135942e-01
## major_categoryHumanities & Liberal Arts       -1.79763232 7.415704e-02
## major_categoryIndustrial Arts & Consumer Services -0.54887249 5.838727e-01
## major_categoryInterdisciplinary               -1.34347667 1.810563e-01
## major_categoryLaw & Public Policy             -0.91647520 3.608233e-01
## major_categoryPhysical Sciences               -0.61045434 5.424435e-01
```

```
## major_categoryPsychology & Social Work            -0.69213657 4.898739e-01
## major_categorySocial Science                      -0.84973074 3.967687e-01
```

```
summary(fit)$coef[,4]<0.05
```

```
##                                       (Intercept)
##                                              TRUE
##                                major_categoryArts
##                                             FALSE
##              major_categoryBiology & Life Science
##                                             FALSE
##                             major_categoryBusiness
##                                             FALSE
##         major_categoryCommunications & Journalism
##                                             FALSE
##             major_categoryComputers & Mathematics
##                                             FALSE
##                            major_categoryEducation
##                                             FALSE
##                          major_categoryEngineering
##                                             FALSE
##                               major_categoryHealth
##                                             FALSE
##           major_categoryHumanities & Liberal Arts
##                                             FALSE
## major_categoryIndustrial Arts & Consumer Services
##                                             FALSE
##                    major_categoryInterdisciplinary
##                                             FALSE
##                 major_categoryLaw & Public Policy
##                                             FALSE
##                   major_categoryPhysical Sciences
##                                             FALSE
##            major_categoryPsychology & Social Work
##                                             FALSE
##                       major_categorySocial Science
##                                             FALSE
```

#the p values suggest that the medians are not significantly different.

```
dance_save("college_major_analysis.rds")
dance_remove()
```