College Major Analysis

Regression Models:

The goal of the course is use regression analysis as tool to understand your data.

To achieve this objective, the course gives the tools to understand regressions, from the simple and practical linear model, to the glm, focussing on binary outcomes (logit) and count data (poisson).

Optional Data Analysis Project:

This analysis is part of the optional quiz, to understand how students understand and work with data.

A codebook for the dataset is given below:

1. rank: Rank by median earnings
2. major_code: Major code
3. major: Major description
4. major_category: Category of major
5. total: Total number of people with major
6. sample_size: Sample size of full-time, year-round individuals used for income/earnings estimates: p25th, median, p75th
7. p25th: 25th percentile of earnings
8. median: Median earnings of full-time, year-round workers
9. p75th: 75th percentile of earnings
10. perc_men: % men with major (out of total)
11. perc_women: % women with major (out of total)
12. perc_employed: % employed (out of total)
13. perc_employed_fulltime: % employed 35 hours or more (out of employed)
14. perc_employed_parttime: % employed less than 35 hours (out of employed)
15. perc_employed_fulltime_yearround: % employed at least 50 weeks and at least 35 hours (out of employed and full-time)
16. perc_unemployed: % unemployed (out of employed)
17. perc_college_jobs: % with job requiring a college degree (out of employed)
18. perc_non_college_jobs: % with job not requiring a college degree (out of employed)
19. perc_low_wage_jobs: % in low-wage service jobs (out of total)

Data Analysis Process Starts:

1. Installing required packages.

```
 1.1 To get started, start a new R/RStudio session with a clean workspace. To do this in R, you can use the q() functi
on to quit, then reopen R. The easiest way to do this in RStudio is to quit RStudio entirely and reopen it. After you
have started a new session, run the following commands. This will load a data.frame called college for you to work wit
h.
```

```
#This will load a data.frame called college for you to work with.
#install.packages("devtools")
#devtools::install_github("jhudsl/collegeIncome", force = TRUE)
library(collegeIncome)
data(college)

#Next download and install the matahari R package with the following commands.

#devtools::install_github("jhudsl/matahari", force = TRUE)
library(matahari)
```

```
    1.2 This package allows a record of your analysis (your R command history) to be documented. You will be uploading a fil
e containing this record to GitHub and submitting the link as part of this quiz.

    1.3 Before you start the analysis for this assignment, enter the following command to begin the documentation of your an
alysis:
```

```
dance_start(value = FALSE, contents = FALSE)
```

2. Calling required packages from the library.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.2
```

```
##
## 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(GGally)
```

```
## Warning: package 'GGally' was built under R version 4.0.2
```

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

3. Getting into the Analysis.

```
#Initial phase of data.frame analysis.

dim(college)
```

```
## [1] 173  19
```

```
str(college)
```

```
## '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" "Metallurgical Enginee
ring" "Naval Architecture And Marine 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 23000 ...
##  $ median                        : num  40000 37000 45000 35000 62000 44700 45000 57000 36000 32200 ...
##  $ p75th                         : num  50000 40000 60000 45000 109000 50000 58000 67000 46000 47100 ...
##  $ 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 ...
```

    3.1 How's the distribution of fulltime employment? Is there a gender difference?

```
g <- ggplot(aes(perc_employed_fulltime_yearround, major_category, colour=perc_women), data=college)
g <- g + geom_point(size=2) + labs(x="Percentage Employed Fulltime Yearround", y="Major Category")
g <- g + scale_colour_gradient(name="Percentage of Women")
```

    3.2 Is there a difference in the median of earnings?

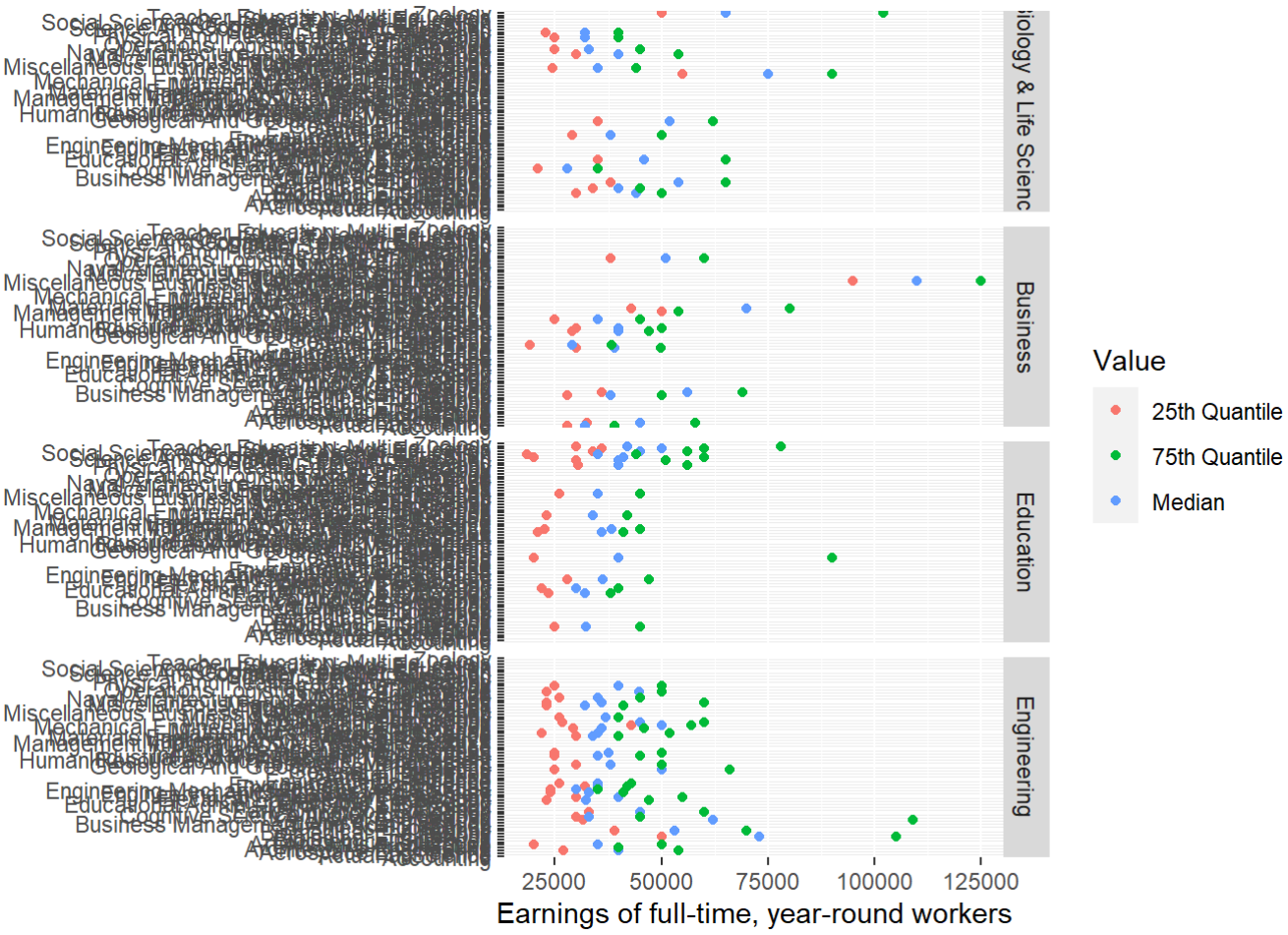
```
h <- ggplot(aes(median, major_category, colour=perc_women), data=college)
h <- h + geom_point(size=2) + labs(x="Median earnings of full-time, year-round workers", y="Major Category") + scale_colour_
gradient(name="Percentage of Women")
ggmatrix(list(g, h), ncol=2, nrow=1, legend=grab_legend(g), xAxisLabels = c("Percentage Employed Fulltime Yearround","Median
earnings of full-time, year-round workers" ))
```

3.3 Looking median, 25th and 75th quintile of earnings.

```
p25th <- data.frame(Earnings = college$p25th, Major_category = college$major_category, Major = college$major, Value="25th Qu
antile")
median <- data.frame(Earnings = college$median, Major_category = college$major_category, Major = college$major, Value="Media
n")
p75th <- data.frame(Earnings = college$p75th, Major_category = college$major_category, Major = college$major, Value="75th Qu
antile")
earnings <- rbind(p25th, median, p75th)
m <- ggplot(aes(Earnings, Major_category), data=earnings)
m <- m + geom_point(aes(colour=Value)) + labs(x="Earnings of full-time, year-round workers", y="Major Category")
m #The higher earnings are Engineering, Education, Business and Biology & Life Science
```

3.4 Investigating these majors further.

```
big4 <- filter(earnings, Major_category=="Engineering"| Major_category=="Education"| Major_category=="Business" | Major_cate
gory=="Biology & Life Science")
l <- ggplot(aes(Earnings, Major), data = big4) + facet_grid(Major_category ~ .)
l <- l + geom_point(aes(colour = Value)) + labs(x="Earnings of full-time, year-round workers", y=NULL)
l
```



3.5 Looking at all the correlations.

```
all <- ggpairs(data = select(college, c(major_category, total, perc_women, median, perc_employed_fulltime:perc_low_wage_jobs)), cardinality_threshold = 17)
all
```

```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).

## Warning: Removed 1 rows containing non-finite values (stat_boxplot).

## Warning: Removed 1 rows containing non-finite values (stat_boxplot).

## Warning: Removed 1 rows containing non-finite values (stat_boxplot).

## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing missing values (geom_text).
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing missing values (geom_text).

## Warning: Removing 1 row that contained a missing value

## Warning: Removing 1 row that contained a missing value

## Warning: Removing 1 row that contained a missing value

## Warning: Removing 1 row that contained a missing value
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing missing values (geom_text).

## Warning: Removing 1 row that contained a missing value

## Warning: Removing 1 row that contained a missing value

## Warning: Removing 1 row that contained a missing value

## Warning: Removing 1 row that contained a missing value
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 1 rows containing non-finite values (stat_density).
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value
```

```
## Warning: Removed 1 rows containing missing values (geom_text).

## Warning: Removed 1 rows containing missing values (geom_text).
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value
```

```
## Warning: Removed 1 rows containing missing values (geom_text).
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value
```

```
## Warning: Removed 1 rows containing missing values (geom_text).
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing non-finite values (stat_density).
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 2 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 2 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 2 rows containing missing values
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removing 1 row that contained a missing value

## Warning: Removing 1 row that contained a missing value

## Warning: Removing 1 row that contained a missing value
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing non-finite values (stat_density).
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing non-finite values (stat_density).
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```

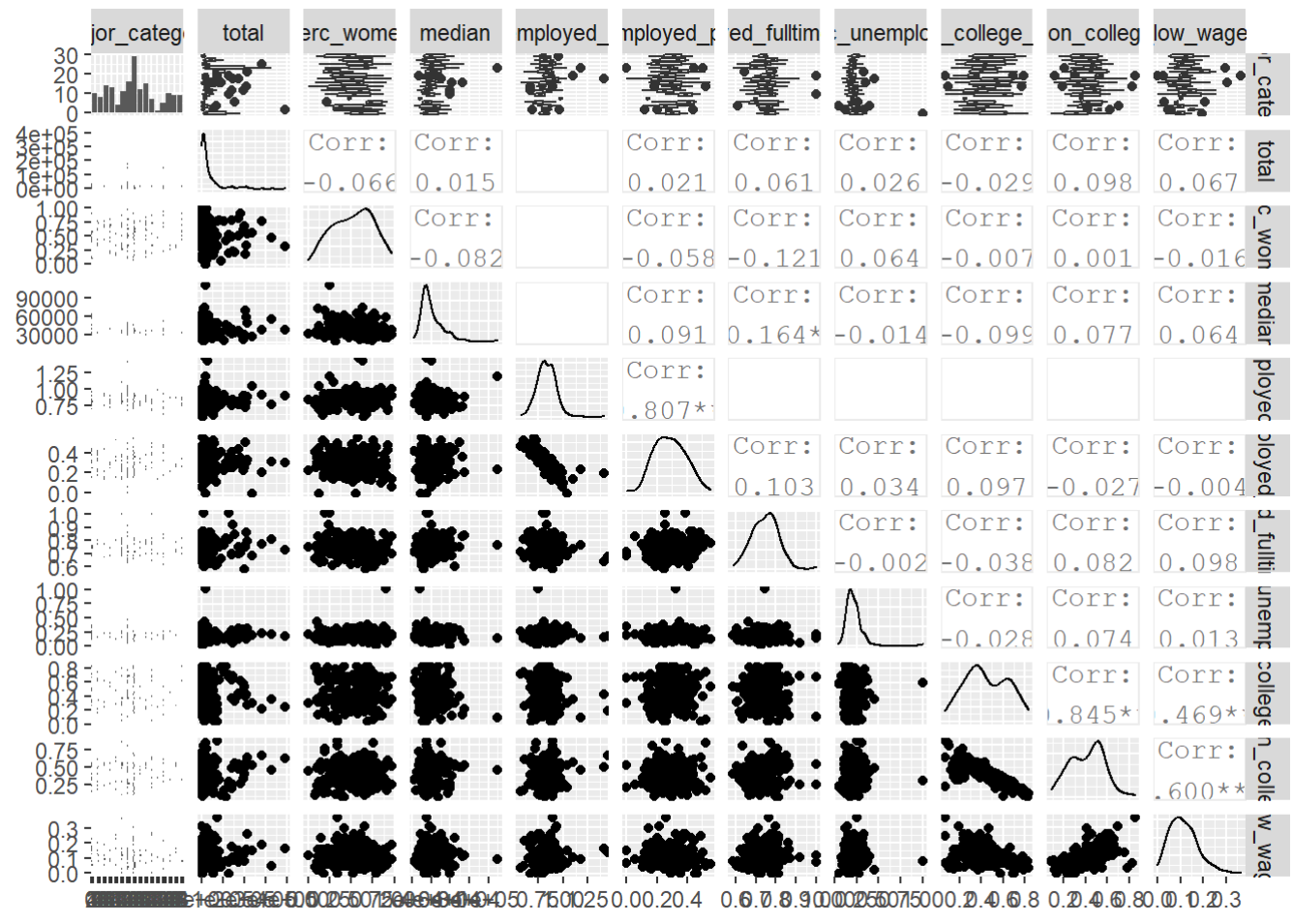
```
## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing non-finite values (stat_density).
```



3.6 Regressions.

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

```
##
## Call:
## lm(formula = median ~ factor(major_category), data = college)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -20154  -6800  -2893   4833  60846
##
## Coefficients:
##                                                 Estimate Std. Error
## (Intercept)                                      43500.0     3590.8
## factor(major_category)Arts                       -5450.0     5386.2
## factor(major_category)Biology & Life Science       364.3     4701.5
## factor(major_category)Business                    5653.8     4776.2
## factor(major_category)Communications & Journalism -1500.0     6717.8
## factor(major_category)Computers & Mathematics     -8781.8     4961.4
## factor(major_category)Education                   -5562.5     4577.4
## factor(major_category)Engineering                 -3106.9     4164.2
## factor(major_category)Health                      -3183.3     4862.0
## factor(major_category)Humanities & Liberal Arts   -8333.3     4635.7
## factor(major_category)Industrial Arts & Consumer Services  -3071.4     5595.9
## factor(major_category)Interdisciplinary          -16000.0    11909.4
## factor(major_category)Law & Public Policy         -5700.0     6219.5
## factor(major_category)Physical Sciences           -3100.0     5078.2
## factor(major_category)Psychology & Social Work    -3611.1     5217.3
## factor(major_category)Social Science              -4433.3     5217.3
##                                                 t value Pr(>|t|)
## (Intercept)                                      12.114   <2e-16 ***
## factor(major_category)Arts                       -1.012   0.3132
## factor(major_category)Biology & Life Science      0.077   0.9383
## factor(major_category)Business                    1.184   0.2383
## factor(major_category)Communications & Journalism -0.223   0.8236
## factor(major_category)Computers & Mathematics     -1.770   0.0787 .
## factor(major_category)Education                   -1.215   0.2261
## factor(major_category)Engineering                 -0.746   0.4567
## factor(major_category)Health                      -0.655   0.5136
## factor(major_category)Humanities & Liberal Arts   -1.798   0.0742 .
## factor(major_category)Industrial Arts & Consumer Services  -0.549   0.5839
## factor(major_category)Interdisciplinary           -1.343   0.1811
## factor(major_category)Law & Public Policy         -0.916   0.3608
## factor(major_category)Physical Sciences           -0.610   0.5424
## factor(major_category)Psychology & Social Work    -0.692   0.4899
## factor(major_category)Social Science              -0.850   0.3968
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11360 on 157 degrees of freedom
## Multiple R-squared:  0.1054, Adjusted R-squared:  0.01995
## F-statistic: 1.233 on 15 and 157 DF,  p-value: 0.2522
```

```
#The major categories weren't significant

fit2 <- lm(median ~ factor(major_category) + perc_women, data = college)
anova(fit, fit2)
```

```
## Analysis of Variance Table
##
## Model 1: median ~ factor(major_category)
## Model 2: median ~ factor(major_category) + perc_women
##   Res.Df        RSS Df Sum of Sq      F Pr(>F)
## 1    157 2.0244e+10
## 2    156 2.0034e+10  1 210042153 1.6356 0.2028
```

```
#Still not able to reject the null hypothesis, thus the regressors are not significant

fit3 <- lm(median ~ perc_women, data = college)
anova(fit3, fit2)
```

```
## Analysis of Variance Table
##
## Model 1: median ~ perc_women
## Model 2: median ~ factor(major_category) + perc_women
##   Res.Df        RSS Df  Sum of Sq      F Pr(>F)
## 1    171 2.2476e+10
## 2    156 2.0034e+10 15 2442945925 1.2682 0.2284
```

```
fit4 <- lm(median ~ factor(major_category)*perc_women, data=college)
summary(fit4)
```

```
##
## Call:
## lm(formula = median ~ factor(major_category) * perc_women, data = college)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -21598  -6157  -2079   5049  36281
##
## Coefficients: (1 not defined because of singularities)
##                                                                 Estimate
## (Intercept)                                                        30716
## factor(major_category)Arts                                          5000
## factor(major_category)Biology & Life Science                       23768
## factor(major_category)Business                                     61626
## factor(major_category)Communications & Journalism                  -3044
## factor(major_category)Computers & Mathematics                       4857
## factor(major_category)Education                                     6696
## factor(major_category)Engineering                                  18145
## factor(major_category)Health                                        7188
## factor(major_category)Humanities & Liberal Arts                     9010
## factor(major_category)Industrial Arts & Consumer Services           6043
## factor(major_category)Interdisciplinary                           -17641
## factor(major_category)Law & Public Policy                           8560
## factor(major_category)Physical Sciences                             3291
## factor(major_category)Psychology & Social Work                     12387
## factor(major_category)Social Science                               17061
## perc_women                                                         30276
## factor(major_category)Arts:perc_women                             -26241
## factor(major_category)Biology & Life Science:perc_women           -47515
## factor(major_category)Business:perc_women                         -105012
## factor(major_category)Communications & Journalism:perc_women       17002
## factor(major_category)Computers & Mathematics:perc_women          -32049
## factor(major_category)Education:perc_women                        -29222
## factor(major_category)Engineering:perc_women                      -46980
## factor(major_category)Health:perc_women                           -24854
## factor(major_category)Humanities & Liberal Arts:perc_women        -38658
## factor(major_category)Industrial Arts & Consumer Services:perc_women -23409
## factor(major_category)Interdisciplinary:perc_women                    NA
## factor(major_category)Law & Public Policy:perc_women              -32447
## factor(major_category)Physical Sciences:perc_women                -16748
## factor(major_category)Psychology & Social Work:perc_women         -36654
## factor(major_category)Social Science:perc_women                   -43674
##                                                                 Std. Error
## (Intercept)                                                           6743
## factor(major_category)Arts                                           15476
## factor(major_category)Biology & Life Science                         10863
## factor(major_category)Business                                       11782
## factor(major_category)Communications & Journalism                    14104
## factor(major_category)Computers & Mathematics                        10329
## factor(major_category)Education                                       9106
## factor(major_category)Engineering                                     8516
## factor(major_category)Health                                          8972
## factor(major_category)Humanities & Liberal Arts                      10244
## factor(major_category)Industrial Arts & Consumer Services            12154
## factor(major_category)Interdisciplinary                              11220
## factor(major_category)Law & Public Policy                            15564
## factor(major_category)Physical Sciences                              10187
## factor(major_category)Psychology & Social Work                       10889
## factor(major_category)Social Science                                 15135
## perc_women                                                           13824
## factor(major_category)Arts:perc_women                                26988
## factor(major_category)Biology & Life Science:perc_women              18995
## factor(major_category)Business:perc_women                            21080
## factor(major_category)Communications & Journalism:perc_women         39395
## factor(major_category)Computers & Mathematics:perc_women             20244
## factor(major_category)Education:perc_women                           17703
## factor(major_category)Engineering:perc_women                         16767
## factor(major_category)Health:perc_women                              17891
## factor(major_category)Humanities & Liberal Arts:perc_women           19143
## factor(major_category)Industrial Arts & Consumer Services:perc_women 22183
## factor(major_category)Interdisciplinary:perc_women                      NA
## factor(major_category)Law & Public Policy:perc_women                 23820
## factor(major_category)Physical Sciences:perc_women                   20030
## factor(major_category)Psychology & Social Work:perc_women            20718
## factor(major_category)Social Science:perc_women                      24405
##                                                                 t value
## (Intercept)                                                       4.555
## factor(major_category)Arts                                        0.323
## factor(major_category)Biology & Life Science                      2.188
## factor(major_category)Business                                    5.231
## factor(major_category)Communications & Journalism                -0.216
## factor(major_category)Computers & Mathematics                     0.470
## factor(major_category)Education                                   0.735
```

```
## factor(major_category)Engineering                                        2.131
## factor(major_category)Health                                             0.801
## factor(major_category)Humanities & Liberal Arts                          0.880
## factor(major_category)Industrial Arts & Consumer Services                0.497
## factor(major_category)Interdisciplinary                                 -1.572
## factor(major_category)Law & Public Policy                                0.550
## factor(major_category)Physical Sciences                                  0.323
## factor(major_category)Psychology & Social Work                           1.137
## factor(major_category)Social Science                                     1.127
## perc_women                                                               2.190
## factor(major_category)Arts:perc_women                                   -0.972
## factor(major_category)Biology & Life Science:perc_women                 -2.501
## factor(major_category)Business:perc_women                               -4.982
## factor(major_category)Communications & Journalism:perc_women             0.432
## factor(major_category)Computers & Mathematics:perc_women                -1.583
## factor(major_category)Education:perc_women                              -1.651
## factor(major_category)Engineering:perc_women                            -2.802
## factor(major_category)Health:perc_women                                 -1.389
## factor(major_category)Humanities & Liberal Arts:perc_women              -2.019
## factor(major_category)Industrial Arts & Consumer Services:perc_women    -1.055
## factor(major_category)Interdisciplinary:perc_women                          NA
## factor(major_category)Law & Public Policy:perc_women                    -1.362
## factor(major_category)Physical Sciences:perc_women                      -0.836
## factor(major_category)Psychology & Social Work:perc_women               -1.769
## factor(major_category)Social Science:perc_women                         -1.790
##                                                                      Pr(>|t|)
## (Intercept)                                                          1.12e-05
## factor(major_category)Arts                                            0.74711
## factor(major_category)Biology & Life Science                          0.03030
## factor(major_category)Business                                       5.93e-07
## factor(major_category)Communications & Journalism                     0.82945
## factor(major_category)Computers & Mathematics                         0.63889
## factor(major_category)Education                                       0.46333
## factor(major_category)Engineering                                     0.03484
## factor(major_category)Health                                          0.42434
## factor(major_category)Humanities & Liberal Arts                       0.38061
## factor(major_category)Industrial Arts & Consumer Services             0.61984
## factor(major_category)Interdisciplinary                               0.11810
## factor(major_category)Law & Public Policy                             0.58318
## factor(major_category)Physical Sciences                               0.74715
## factor(major_category)Psychology & Social Work                        0.25725
## factor(major_category)Social Science                                  0.26155
## perc_women                                                            0.03015
## factor(major_category)Arts:perc_women                                 0.33254
## factor(major_category)Biology & Life Science:perc_women               0.01350
## factor(major_category)Business:perc_women                            1.81e-06
## factor(major_category)Communications & Journalism:perc_women          0.66670
## factor(major_category)Computers & Mathematics:perc_women              0.11562
## factor(major_category)Education:perc_women                            0.10102
## factor(major_category)Engineering:perc_women                          0.00579
## factor(major_category)Health:perc_women                               0.16696
## factor(major_category)Humanities & Liberal Arts:perc_women            0.04533
## factor(major_category)Industrial Arts & Consumer Services:perc_women  0.29309
## factor(major_category)Interdisciplinary:perc_women                        NA
## factor(major_category)Law & Public Policy:perc_women                  0.17530
## factor(major_category)Physical Sciences:perc_women                    0.40450
## factor(major_category)Psychology & Social Work:perc_women             0.07901
## factor(major_category)Social Science:perc_women                       0.07566
##
## (Intercept)                                                          ***
## factor(major_category)Arts
## factor(major_category)Biology & Life Science                         *
## factor(major_category)Business                                       ***
## factor(major_category)Communications & Journalism
## factor(major_category)Computers & Mathematics
## factor(major_category)Education
## factor(major_category)Engineering                                    *
## factor(major_category)Health
## factor(major_category)Humanities & Liberal Arts
## factor(major_category)Industrial Arts & Consumer Services
## factor(major_category)Interdisciplinary
## factor(major_category)Law & Public Policy
## factor(major_category)Physical Sciences
## factor(major_category)Psychology & Social Work
## factor(major_category)Social Science
## perc_women                                                           *
## factor(major_category)Arts:perc_women
## factor(major_category)Biology & Life Science:perc_women              *
## factor(major_category)Business:perc_women                            ***
## factor(major_category)Communications & Journalism:perc_women
## factor(major_category)Computers & Mathematics:perc_women
## factor(major_category)Education:perc_women
## factor(major_category)Engineering:perc_women                         **
```

```
## factor(major_category)Health:perc_women
## factor(major_category)Humanities & Liberal Arts:perc_women           *
## factor(major_category)Industrial Arts & Consumer Services:perc_women
## factor(major_category)Interdisciplinary:perc_women
## factor(major_category)Law & Public Policy:perc_women
## factor(major_category)Physical Sciences:perc_women
## factor(major_category)Psychology & Social Work:perc_women            .
## factor(major_category)Social Science:perc_women                      .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10670 on 142 degrees of freedom
## Multiple R-squared:  0.2851, Adjusted R-squared:  0.134
## F-statistic: 1.887 on 30 and 142 DF,  p-value: 0.007361
```

```
#Now there are some very significant coef
anova(fit, fit4)
```

```
## Analysis of Variance Table
##
## Model 1: median ~ factor(major_category)
## Model 2: median ~ factor(major_category) * perc_women
##   Res.Df        RSS Df  Sum of Sq      F   Pr(>F)
## 1    157 2.0244e+10
## 2    142 1.6178e+10 15 4065245219 2.3788 0.004338 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#The inclusion of the interation between the percentage of woman and the major category

anova(fit3, fit4)
```

```
## Analysis of Variance Table
##
## Model 1: median ~ perc_women
## Model 2: median ~ factor(major_category) * perc_women
##   Res.Df        RSS Df  Sum of Sq      F   Pr(>F)
## 1    171 2.2476e+10
## 2    142 1.6178e+10 29 6298148991 1.9062 0.007153 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#was able to reject the null hypothesis, thus indicating the significance
#Still, the R squared is low, even though the large number of variables, as shown by the adjusted R-squared

fit5 <- lm(p25th ~ factor(major_category)*perc_women, data=college)
summary(fit5)
```

```
##
## Call:
## lm(formula = p25th ~ factor(major_category) * perc_women, data = college)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -19033  -5537  -1719   4398  40358
##
## Coefficients: (1 not defined because of singularities)
##                                                                 Estimate
## (Intercept)                                                      24554.2
## factor(major_category)Arts                                        9296.8
## factor(major_category)Biology & Life Science                     13825.1
## factor(major_category)Business                                   43305.8
## factor(major_category)Communications & Journalism                -3777.3
## factor(major_category)Computers & Mathematics                     2886.1
## factor(major_category)Education                                  -3307.5
## factor(major_category)Engineering                                 8280.4
## factor(major_category)Health                                     -1616.6
## factor(major_category)Humanities & Liberal Arts                   2438.5
## factor(major_category)Industrial Arts & Consumer Services         2218.0
## factor(major_category)Interdisciplinary                          -8137.7
## factor(major_category)Law & Public Policy                         3208.8
## factor(major_category)Physical Sciences                           737.1
## factor(major_category)Psychology & Social Work                    9578.5
## factor(major_category)Social Science                             13060.7
## perc_women                                                       13607.5
## factor(major_category)Arts:perc_women                           -22381.9
## factor(major_category)Biology & Life Science:perc_women         -23221.4
## factor(major_category)Business:perc_women                       -66650.7
## factor(major_category)Communications & Journalism:perc_women     30025.3
## factor(major_category)Computers & Mathematics:perc_women        -15463.2
## factor(major_category)Education:perc_women                       -4835.8
## factor(major_category)Engineering:perc_women                    -21838.3
## factor(major_category)Health:perc_women                           -168.3
## factor(major_category)Humanities & Liberal Arts:perc_women      -15493.8
## factor(major_category)Industrial Arts & Consumer Services:perc_women  -6766.6
## factor(major_category)Interdisciplinary:perc_women                    NA
## factor(major_category)Law & Public Policy:perc_women            -13847.2
## factor(major_category)Physical Sciences:perc_women                -935.1
## factor(major_category)Psychology & Social Work:perc_women       -19692.6
## factor(major_category)Social Science:perc_women                 -26928.2
##                                                                 Std. Error
## (Intercept)                                                         5479.6
## factor(major_category)Arts                                         12576.0
## factor(major_category)Biology & Life Science                        8827.4
## factor(major_category)Business                                      9574.2
## factor(major_category)Communications & Journalism                  11461.7
## factor(major_category)Computers & Mathematics                       8393.8
## factor(major_category)Education                                     7400.1
## factor(major_category)Engineering                                   6920.6
## factor(major_category)Health                                        7290.7
## factor(major_category)Humanities & Liberal Arts                     8324.9
## factor(major_category)Industrial Arts & Consumer Services           9876.9
## factor(major_category)Interdisciplinary                             9117.7
## factor(major_category)Law & Public Policy                          12648.0
## factor(major_category)Physical Sciences                             8278.6
## factor(major_category)Psychology & Social Work                      8849.1
## factor(major_category)Social Science                               12299.1
## perc_women                                                         11234.1
## factor(major_category)Arts:perc_women                              21931.0
## factor(major_category)Biology & Life Science:perc_women            15435.9
## factor(major_category)Business:perc_women                          17130.6
## factor(major_category)Communications & Journalism:perc_women       32013.7
## factor(major_category)Computers & Mathematics:perc_women           16451.1
## factor(major_category)Education:perc_women                         14386.2
## factor(major_category)Engineering:perc_women                       13625.0
## factor(major_category)Health:perc_women                            14538.8
## factor(major_category)Humanities & Liberal Arts:perc_women         15556.3
## factor(major_category)Industrial Arts & Consumer Services:perc_women  18026.9
## factor(major_category)Interdisciplinary:perc_women                     NA
## factor(major_category)Law & Public Policy:perc_women               19356.7
## factor(major_category)Physical Sciences:perc_women                 16277.2
## factor(major_category)Psychology & Social Work:perc_women          16836.1
## factor(major_category)Social Science:perc_women                    19832.6
##                                                                t value
## (Intercept)                                                      4.481
## factor(major_category)Arts                                       0.739
## factor(major_category)Biology & Life Science                     1.566
## factor(major_category)Business                                   4.523
## factor(major_category)Communications & Journalism               -0.330
## factor(major_category)Computers & Mathematics                    0.344
## factor(major_category)Education                                 -0.447
```

```
## factor(major_category)Engineering                                      1.196
## factor(major_category)Health                                          -0.222
## factor(major_category)Humanities & Liberal Arts                        0.293
## factor(major_category)Industrial Arts & Consumer Services              0.225
## factor(major_category)Interdisciplinary                               -0.893
## factor(major_category)Law & Public Policy                              0.254
## factor(major_category)Physical Sciences                                0.089
## factor(major_category)Psychology & Social Work                         1.082
## factor(major_category)Social Science                                   1.062
## perc_women                                                             1.211
## factor(major_category)Arts:perc_women                                 -1.021
## factor(major_category)Biology & Life Science:perc_women               -1.504
## factor(major_category)Business:perc_women                             -3.891
## factor(major_category)Communications & Journalism:perc_women           0.938
## factor(major_category)Computers & Mathematics:perc_women              -0.940
## factor(major_category)Education:perc_women                            -0.336
## factor(major_category)Engineering:perc_women                          -1.603
## factor(major_category)Health:perc_women                               -0.012
## factor(major_category)Humanities & Liberal Arts:perc_women            -0.996
## factor(major_category)Industrial Arts & Consumer Services:perc_women  -0.375
## factor(major_category)Interdisciplinary:perc_women                        NA
## factor(major_category)Law & Public Policy:perc_women                  -0.715
## factor(major_category)Physical Sciences:perc_women                    -0.057
## factor(major_category)Psychology & Social Work:perc_women             -1.170
## factor(major_category)Social Science:perc_women                       -1.358
##                                                                     Pr(>|t|)
## (Intercept)                                                         1.52e-05
## factor(major_category)Arts                                          0.460978
## factor(major_category)Biology & Life Science                        0.119536
## factor(major_category)Business                                      1.28e-05
## factor(major_category)Communications & Journalism                   0.742221
## factor(major_category)Computers & Mathematics                       0.731474
## factor(major_category)Education                                     0.655594
## factor(major_category)Engineering                                   0.233500
## factor(major_category)Health                                        0.824835
## factor(major_category)Humanities & Liberal Arts                     0.770013
## factor(major_category)Industrial Arts & Consumer Services           0.822638
## factor(major_category)Interdisciplinary                             0.373627
## factor(major_category)Law & Public Policy                           0.800096
## factor(major_category)Physical Sciences                             0.929179
## factor(major_category)Psychology & Social Work                      0.280898
## factor(major_category)Social Science                                0.290075
## perc_women                                                          0.227805
## factor(major_category)Arts:perc_women                               0.309199
## factor(major_category)Biology & Life Science:perc_women             0.134706
## factor(major_category)Business:perc_women                           0.000153
## factor(major_category)Communications & Journalism:perc_women        0.349894
## factor(major_category)Computers & Mathematics:perc_women            0.348840
## factor(major_category)Education:perc_women                          0.737260
## factor(major_category)Engineering:perc_women                        0.111198
## factor(major_category)Health:perc_women                             0.990778
## factor(major_category)Humanities & Liberal Arts:perc_women          0.320954
## factor(major_category)Industrial Arts & Consumer Services:perc_women 0.707952
## factor(major_category)Interdisciplinary:perc_women                        NA
## factor(major_category)Law & Public Policy:perc_women                0.475556
## factor(major_category)Physical Sciences:perc_women                  0.954270
## factor(major_category)Psychology & Social Work:perc_women           0.244094
## factor(major_category)Social Science:perc_women                     0.176688
##
## (Intercept)                                                         ***
## factor(major_category)Arts
## factor(major_category)Biology & Life Science
## factor(major_category)Business                                      ***
## factor(major_category)Communications & Journalism
## factor(major_category)Computers & Mathematics
## factor(major_category)Education
## factor(major_category)Engineering
## factor(major_category)Health
## factor(major_category)Humanities & Liberal Arts
## factor(major_category)Industrial Arts & Consumer Services
## factor(major_category)Interdisciplinary
## factor(major_category)Law & Public Policy
## factor(major_category)Physical Sciences
## factor(major_category)Psychology & Social Work
## factor(major_category)Social Science
## perc_women
## factor(major_category)Arts:perc_women
## factor(major_category)Biology & Life Science:perc_women
## factor(major_category)Business:perc_women                           ***
## factor(major_category)Communications & Journalism:perc_women
## factor(major_category)Computers & Mathematics:perc_women
## factor(major_category)Education:perc_women
## factor(major_category)Engineering:perc_women
```

```
## factor(major_category)Health:perc_women
## factor(major_category)Humanities & Liberal Arts:perc_women
## factor(major_category)Industrial Arts & Consumer Services:perc_women
## factor(major_category)Interdisciplinary:perc_women
## factor(major_category)Law & Public Policy:perc_women
## factor(major_category)Physical Sciences:perc_women
## factor(major_category)Psychology & Social Work:perc_women
## factor(major_category)Social Science:perc_women
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8674 on 142 degrees of freedom
## Multiple R-squared:  0.2607, Adjusted R-squared:  0.1045
## F-statistic: 1.669 on 30 and 142 DF,  p-value: 0.02524
```

```
#The results for the model using the 25th quintile don't hold as well as the model for the median earnings
#The R^2 is lower, and it's unable to reject the null hypothesis for the F-statistic

fit6 <- lm(p75th ~ factor(major_category)*perc_women, data=college)
summary(fit6)
```

```
##
## Call:
## lm(formula = p75th ~ factor(major_category) * perc_women, data = college)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -26309  -9145  -3476   5973  53199
##
## Coefficients: (1 not defined because of singularities)
##                                                                 Estimate
## (Intercept)                                                        39902
## factor(major_category)Arts                                          1717
## factor(major_category)Biology & Life Science                       28374
## factor(major_category)Business                                     65239
## factor(major_category)Communications & Journalism                  -2670
## factor(major_category)Computers & Mathematics                       8645
## factor(major_category)Education                                    18475
## factor(major_category)Engineering                                  27270
## factor(major_category)Health                                       10213
## factor(major_category)Humanities & Liberal Arts                     6643
## factor(major_category)Industrial Arts & Consumer Services           8886
## factor(major_category)Interdisciplinary                           -17359
## factor(major_category)Law & Public Policy                          13724
## factor(major_category)Physical Sciences                             2854
## factor(major_category)Psychology & Social Work                     19236
## factor(major_category)Social Science                               28332
## perc_women                                                         32440
## factor(major_category)Arts:perc_women                             -26142
## factor(major_category)Biology & Life Science:perc_women           -52020
## factor(major_category)Business:perc_women                         -112526
## factor(major_category)Communications & Journalism:perc_women        9690
## factor(major_category)Computers & Mathematics:perc_women          -40926
## factor(major_category)Education:perc_women                        -44491
## factor(major_category)Engineering:perc_women                      -59647
## factor(major_category)Health:perc_women                           -31013
## factor(major_category)Humanities & Liberal Arts:perc_women        -33197
## factor(major_category)Industrial Arts & Consumer Services:perc_women -28249
## factor(major_category)Interdisciplinary:perc_women                    NA
## factor(major_category)Law & Public Policy:perc_women              -38655
## factor(major_category)Physical Sciences:perc_women                -15206
## factor(major_category)Psychology & Social Work:perc_women         -48813
## factor(major_category)Social Science:perc_women                   -56558
##                                                                 Std. Error
## (Intercept)                                                           9183
## factor(major_category)Arts                                           21076
## factor(major_category)Biology & Life Science                         14794
## factor(major_category)Business                                       16045
## factor(major_category)Communications & Journalism                    19208
## factor(major_category)Computers & Mathematics                        14067
## factor(major_category)Education                                      12402
## factor(major_category)Engineering                                    11598
## factor(major_category)Health                                         12218
## factor(major_category)Humanities & Liberal Arts                      13951
## factor(major_category)Industrial Arts & Consumer Services            16552
## factor(major_category)Interdisciplinary                              15280
## factor(major_category)Law & Public Policy                            21196
## factor(major_category)Physical Sciences                              13874
## factor(major_category)Psychology & Social Work                       14830
## factor(major_category)Social Science                                 20612
## perc_women                                                           18827
## factor(major_category)Arts:perc_women                                36753
## factor(major_category)Biology & Life Science:perc_women              25868
## factor(major_category)Business:perc_women                            28708
## factor(major_category)Communications & Journalism:perc_women         53651
## factor(major_category)Computers & Mathematics:perc_women             27570
## factor(major_category)Education:perc_women                           24109
## factor(major_category)Engineering:perc_women                         22834
## factor(major_category)Health:perc_women                              24365
## factor(major_category)Humanities & Liberal Arts:perc_women           26070
## factor(major_category)Industrial Arts & Consumer Services:perc_women 30211
## factor(major_category)Interdisciplinary:perc_women                      NA
## factor(major_category)Law & Public Policy:perc_women                 32439
## factor(major_category)Physical Sciences:perc_women                   27278
## factor(major_category)Psychology & Social Work:perc_women            28215
## factor(major_category)Social Science:perc_women                      33237
##                                                                 t value
## (Intercept)                                                       4.345
## factor(major_category)Arts                                        0.081
## factor(major_category)Biology & Life Science                      1.918
## factor(major_category)Business                                    4.066
## factor(major_category)Communications & Journalism                -0.139
## factor(major_category)Computers & Mathematics                     0.615
## factor(major_category)Education                                   1.490
```

```
## factor(major_category)Engineering                                        2.351
## factor(major_category)Health                                             0.836
## factor(major_category)Humanities & Liberal Arts                          0.476
## factor(major_category)Industrial Arts & Consumer Services                0.537
## factor(major_category)Interdisciplinary                                 -1.136
## factor(major_category)Law & Public Policy                                0.647
## factor(major_category)Physical Sciences                                  0.206
## factor(major_category)Psychology & Social Work                           1.297
## factor(major_category)Social Science                                     1.375
## perc_women                                                               1.723
## factor(major_category)Arts:perc_women                                   -0.711
## factor(major_category)Biology & Life Science:perc_women                 -2.011
## factor(major_category)Business:perc_women                               -3.920
## factor(major_category)Communications & Journalism:perc_women             0.181
## factor(major_category)Computers & Mathematics:perc_women                -1.484
## factor(major_category)Education:perc_women                              -1.845
## factor(major_category)Engineering:perc_women                            -2.612
## factor(major_category)Health:perc_women                                 -1.273
## factor(major_category)Humanities & Liberal Arts:perc_women              -1.273
## factor(major_category)Industrial Arts & Consumer Services:perc_women    -0.935
## factor(major_category)Interdisciplinary:perc_women                          NA
## factor(major_category)Law & Public Policy:perc_women                    -1.192
## factor(major_category)Physical Sciences:perc_women                      -0.557
## factor(major_category)Psychology & Social Work:perc_women               -1.730
## factor(major_category)Social Science:perc_women                         -1.702
##                                                                       Pr(>|t|)
## (Intercept)                                                           2.63e-05
## factor(major_category)Arts                                            0.935177
## factor(major_category)Biology & Life Science                          0.057121
## factor(major_category)Business                                        7.89e-05
## factor(major_category)Communications & Journalism                     0.889655
## factor(major_category)Computers & Mathematics                         0.539837
## factor(major_category)Education                                       0.138506
## factor(major_category)Engineering                                     0.020085
## factor(major_category)Health                                          0.404628
## factor(major_category)Humanities & Liberal Arts                       0.634693
## factor(major_category)Industrial Arts & Consumer Services             0.592211
## factor(major_category)Interdisciplinary                               0.257854
## factor(major_category)Law & Public Policy                             0.518374
## factor(major_category)Physical Sciences                               0.837329
## factor(major_category)Psychology & Social Work                        0.196690
## factor(major_category)Social Science                                  0.171438
## perc_women                                                            0.087051
## factor(major_category)Arts:perc_women                                 0.478070
## factor(major_category)Biology & Life Science:perc_women               0.046223
## factor(major_category)Business:perc_women                             0.000137
## factor(major_category)Communications & Journalism:perc_women          0.856931
## factor(major_category)Computers & Mathematics:perc_women              0.139905
## factor(major_category)Education:perc_women                            0.067061
## factor(major_category)Engineering:perc_women                          0.009962
## factor(major_category)Health:perc_women                               0.205154
## factor(major_category)Humanities & Liberal Arts:perc_women            0.204964
## factor(major_category)Industrial Arts & Consumer Services:perc_women  0.351347
## factor(major_category)Interdisciplinary:perc_women                          NA
## factor(major_category)Law & Public Policy:perc_women                  0.235403
## factor(major_category)Physical Sciences:perc_women                    0.578093
## factor(major_category)Psychology & Social Work:perc_women             0.085797
## factor(major_category)Social Science:perc_women                       0.091005
##
## (Intercept)                                                           ***
## factor(major_category)Arts
## factor(major_category)Biology & Life Science                          .
## factor(major_category)Business                                        ***
## factor(major_category)Communications & Journalism
## factor(major_category)Computers & Mathematics
## factor(major_category)Education
## factor(major_category)Engineering                                     *
## factor(major_category)Health
## factor(major_category)Humanities & Liberal Arts
## factor(major_category)Industrial Arts & Consumer Services
## factor(major_category)Interdisciplinary
## factor(major_category)Law & Public Policy
## factor(major_category)Physical Sciences
## factor(major_category)Psychology & Social Work
## factor(major_category)Social Science
## perc_women                                                            .
## factor(major_category)Arts:perc_women
## factor(major_category)Biology & Life Science:perc_women               *
## factor(major_category)Business:perc_women                             ***
## factor(major_category)Communications & Journalism:perc_women
## factor(major_category)Computers & Mathematics:perc_women
## factor(major_category)Education:perc_women                            .
## factor(major_category)Engineering:perc_women                          **
```

```
## factor(major_category)Health:perc_women
## factor(major_category)Humanities & Liberal Arts:perc_women
## factor(major_category)Industrial Arts & Consumer Services:perc_women
## factor(major_category)Interdisciplinary:perc_women
## factor(major_category)Law & Public Policy:perc_women
## factor(major_category)Physical Sciences:perc_women
## factor(major_category)Psychology & Social Work:perc_women            .
## factor(major_category)Social Science:perc_women                      .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14540 on 142 degrees of freedom
## Multiple R-squared:  0.2149, Adjusted R-squared:  0.04903
## F-statistic: 1.296 on 30 and 142 DF,  p-value: 0.1599
```

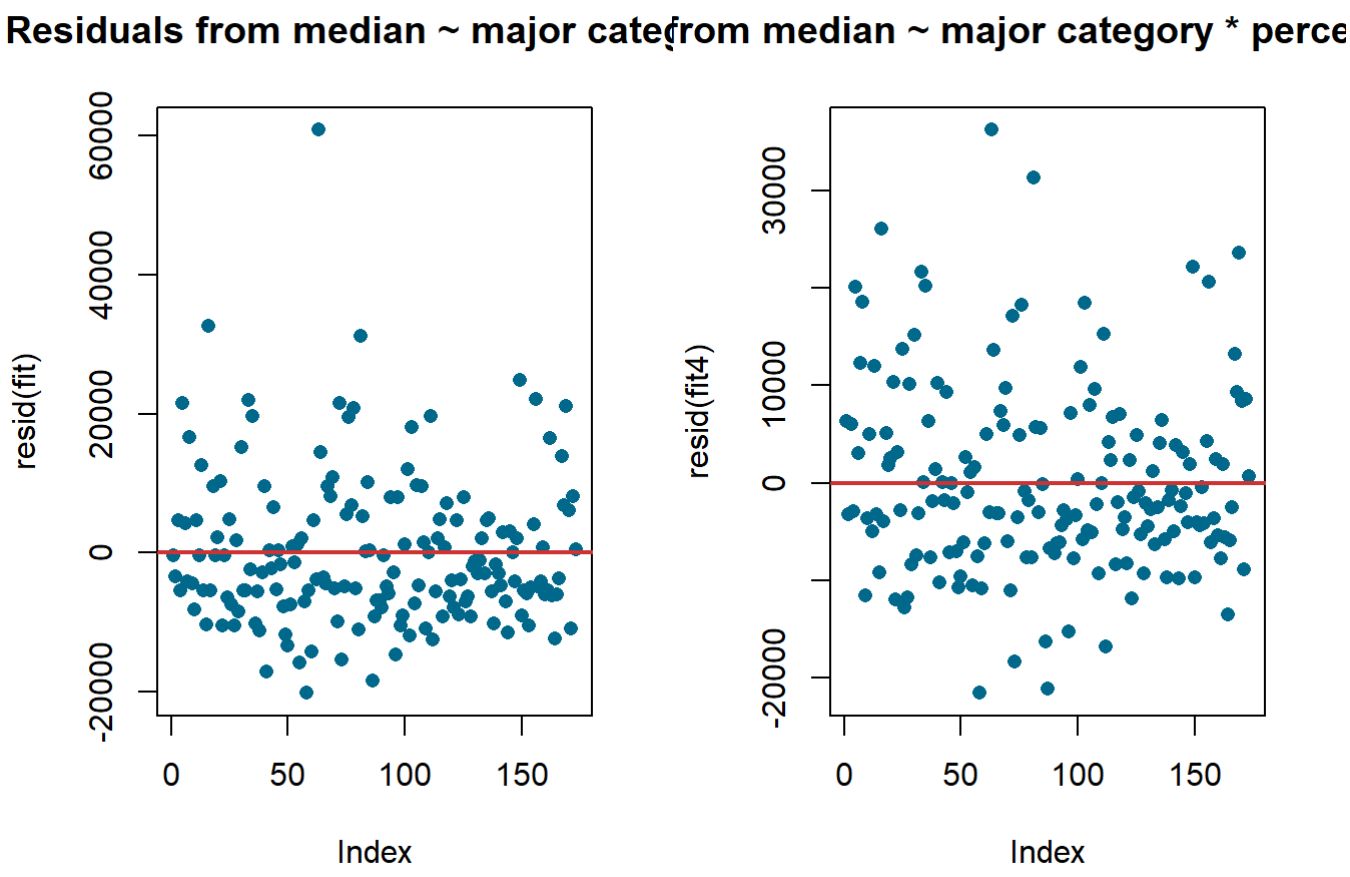
```
#The model for the 75th quintile has a smaller adjusted p-value
```

```
   3.7 Conclusions.
```

- An interesting conclusion is that for 3 out of the 4 major categories investigated further,
- Education, Engineering, Business and Biology & Life Science, because of the high earnings,their interations with the percentage of women in these areas were significant and their coefficients were negative. Looking at the graphic “h” it’s possible to visualize this effect.
- Even though this major categories appear to receive higher earnings, the greater the female participation in the area,the lower their gains.

```
  3.8 Looking at the residuals.
```

```
par(mfrow=c(1,2))
plot(resid(fit), pch=16, col="deepskyblue4")
title("Residuals from median ~ major category")
abline(h=0, col="brown3", lwd=2)
plot(resid(fit4), pch=16, col="deepskyblue4")
abline(h=0, col="brown3", lwd=2)
title("Residuals from median ~ major category * percetage of women")
```




```
#The residuals for the model with the interation are closer to zero, and appear random enough.

fit7 <- lm(median ~ major_category*perc_women + perc_college_jobs + perc_employed + perc_low_wage_jobs, data = college)
summary(fit7)
```

```
##
## Call:
## lm(formula = median ~ major_category * perc_women + perc_college_jobs +
##     perc_employed + perc_low_wage_jobs, data = college)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -21866  -6246  -2160   5095  35045
##
## Coefficients: (1 not defined because of singularities)
##                                                        Estimate
## (Intercept)                                               29509
## major_categoryArts                                         2663
## major_categoryBiology & Life Science                      23291
## major_categoryBusiness                                    59641
## major_categoryCommunications & Journalism                 -3503
## major_categoryComputers & Mathematics                      2740
## major_categoryEducation                                    6585
## major_categoryEngineering                                 17207
## major_categoryHealth                                       6485
## major_categoryHumanities & Liberal Arts                    7555
## major_categoryIndustrial Arts & Consumer Services          5366
## major_categoryInterdisciplinary                          -19355
## major_categoryLaw & Public Policy                          7450
## major_categoryPhysical Sciences                            2729
## major_categoryPsychology & Social Work                    10505
## major_categorySocial Science                              14878
## perc_women                                                29631
## perc_college_jobs                                         -4976
## perc_employed                                              5773
## perc_low_wage_jobs                                        -2039
## major_categoryArts:perc_women                             -23315
## major_categoryBiology & Life Science:perc_women           -47539
## major_categoryBusiness:perc_women                        -102874
## major_categoryCommunications & Journalism:perc_women      16332
## major_categoryComputers & Mathematics:perc_women          -29311
## major_categoryEducation:perc_women                        -29455
## major_categoryEngineering:perc_women                      -46183
## major_categoryHealth:perc_women                           -24614
## major_categoryHumanities & Liberal Arts:perc_women        -37588
## major_categoryIndustrial Arts & Consumer Services:perc_women -22626
## major_categoryInterdisciplinary:perc_women                    NA
## major_categoryLaw & Public Policy:perc_women              -32312
## major_categoryPhysical Sciences:perc_women                -16539
## major_categoryPsychology & Social Work:perc_women         -35516
## major_categorySocial Science:perc_women                   -39389
##                                                       Std. Error t value
## (Intercept)                                                11770   2.507
## major_categoryArts                                         15920   0.167
## major_categoryBiology & Life Science                       10989   2.120
## major_categoryBusiness                                     12115   4.923
## major_categoryCommunications & Journalism                  14334  -0.244
## major_categoryComputers & Mathematics                      10723   0.256
## major_categoryEducation                                     9320   0.707
## major_categoryEngineering                                   8647   1.990
## major_categoryHealth                                        9080   0.714
## major_categoryHumanities & Liberal Arts                    10549   0.716
## major_categoryIndustrial Arts & Consumer Services          12324   0.435
## major_categoryInterdisciplinary                            11447  -1.691
## major_categoryLaw & Public Policy                          15823   0.471
## major_categoryPhysical Sciences                            10327   0.264
## major_categoryPsychology & Social Work                     11192   0.939
## major_categorySocial Science                               15874   0.937
## perc_women                                                 14124   2.098
## perc_college_jobs                                           5166  -0.963
## perc_employed                                              10569   0.546
## perc_low_wage_jobs                                         15837  -0.129
## major_categoryArts:perc_women                              27807  -0.838
## major_categoryBiology & Life Science:perc_women            19239  -2.471
## major_categoryBusiness:perc_women                          21674  -4.746
## major_categoryCommunications & Journalism:perc_women       40290   0.405
## major_categoryComputers & Mathematics:perc_women           21408  -1.369
## major_categoryEducation:perc_women                         18254  -1.614
## major_categoryEngineering:perc_women                       17062  -2.707
## major_categoryHealth:perc_women                            18102  -1.360
## major_categoryHumanities & Liberal Arts:perc_women         19644  -1.913
## major_categoryIndustrial Arts & Consumer Services:perc_women 22601  -1.001
## major_categoryInterdisciplinary:perc_women                    NA      NA
## major_categoryLaw & Public Policy:perc_women               24306  -1.329
## major_categoryPhysical Sciences:perc_women                 20382  -0.811
## major_categoryPsychology & Social Work:perc_women          21086  -1.684
## major_categorySocial Science:perc_women                    25667  -1.535
##                                                      Pr(>|t|)
```

```
## (Intercept)                                          0.01333 *
## major_categoryArts                                   0.86738
## major_categoryBiology & Life Science                 0.03584 *
## major_categoryBusiness                              2.40e-06 ***
## major_categoryCommunications & Journalism            0.80728
## major_categoryComputers & Mathematics                0.79868
## major_categoryEducation                              0.48104
## major_categoryEngineering                            0.04859 *
## major_categoryHealth                                 0.47632
## major_categoryHumanities & Liberal Arts              0.47509
## major_categoryIndustrial Arts & Consumer Services    0.66397
## major_categoryInterdisciplinary                      0.09314 .
## major_categoryLaw & Public Policy                    0.63848
## major_categoryPhysical Sciences                      0.79196
## major_categoryPsychology & Social Work               0.34956
## major_categorySocial Science                         0.35027
## perc_women                                           0.03773 *
## perc_college_jobs                                    0.33718
## perc_employed                                        0.58580
## perc_low_wage_jobs                                   0.89775
## major_categoryArts:perc_women                        0.40323
## major_categoryBiology & Life Science:perc_women      0.01469 *
## major_categoryBusiness:perc_women                   5.11e-06 ***
## major_categoryCommunications & Journalism:perc_women 0.68584
## major_categoryComputers & Mathematics:perc_women     0.17317
## major_categoryEducation:perc_women                   0.10888
## major_categoryEngineering:perc_women                 0.00765 **
## major_categoryHealth:perc_women                      0.17613
## major_categoryHumanities & Liberal Arts:perc_women   0.05776 .
## major_categoryIndustrial Arts & Consumer Services:perc_women  0.31852
## major_categoryInterdisciplinary:perc_women                NA
## major_categoryLaw & Public Policy:perc_women         0.18591
## major_categoryPhysical Sciences:perc_women           0.41849
## major_categoryPsychology & Social Work:perc_women    0.09437 .
## major_categorySocial Science:perc_women              0.12716
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10770 on 138 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.2902, Adjusted R-squared:  0.1205
## F-statistic:  1.71 on 33 and 138 DF,  p-value: 0.0174
```

```
#When you have finished your analysis, use the following command to save the record of your analysis on your desktop:

dance_save("college_major_analysis.rds")
```

Conclusion:

- Including other variables doesn't bring much to the model, their coefficients are not significant and the R^2 (and adjusted R^2) and the F-statistic are slightly worse.
- The best model was the forth, with the interaction between the percentage of women and the major category.
- Some interesting coefficients are significant and the model and its F-statistic is able to reject the null hypothesis that all the coefficients are zero. But still, the effect is minimal, and the model doesn't explain much of the outcome.
- The first model, with only the major categories as regressors indicate that it does not affect the earnings.