GENDER, MAJORS, AND SALARY GAP

NAME

DATE

In fulfillment of the requirements for STAT 234

Professor Priya Kohli

Abstract

In this paper, we explored the gender role in predicting salary in different majors. We saw significant effects of gender and the choice of major on salary, but we did not see the interaction term of these two factors. The result confirms our hypothesis that there is another third variable influencing both gender and the choice of major.

Introduction

Education is frequently referred to as a human capital investment. People invest in human capital for the same reasons they invest in financial assets, which include the desire to make profits. Amongst factors (social-economic background, gender, race, and wealth status), college choice does significantly correlate with the expected earnings if students finish in a chosen field (Montmarquette et al., 2002). The relationship between salary and major also vary in terms of gender and race. More specifically, the research showed women are less influenced by major compared to men and non-whites more than whites (Montmarquette et al., 2002 & Dowell, 2022). We chose datasets 2010-2012 and 2013-2019 American Community Survey because they have detailed earnings of each major and gender.

Through initial data analysis, we learned that men-dominated majors tend to earn more than women-dominated majors (Premise 1)(Figure 1). Men tend to major more in STEM (Premise 2)(Figure 1). These majors tend to be in the top earnings majors (Figure 3). From these two premises, we have 3 possible scenarios. The first scenario is that men earn more because men are men (gender effect) (1). The second scenario is that men earn more because STEM majors tend to earn more (2). The third scenario is that men earn more because there is a third variable that influences both gender and engineering and math majors (3). If (1) is true, then we will see men are consistently paid more across majors. If (2) is true, then we will see engineering and math majors are paid more, compared to other majors. If (3) is true, then we will see that both gender and majors have a significant impact on salary or the interaction term between gender and majors have a significant effect on salary. Based on other published papers, we hypothesize that both gender and the choice of major have a significant effect on the salary range (3 is true).

Statistical Analysis with Data Visualization

In this analysis, we use multivariate graphs to explore the question. We use a line plot to show how differences between men's and women's salaries within the same occupation. In this plot, people can see how this disparity is shortened/lengthened depending on the top, middle, and low earning groups. We also create a bar plot, demonstrating the difference in salary from 2013 to 2016. Each plot has an overlaying line graph, showing the increase/decrease and general trend over time. We help the readers understand the context better by dividing it into 3 earning groups. So each graph will involve 4 levels of information. These techniques combine both categorical and nominal variables in the same graph while incorporating 2 factors (Gender and Group) into the graph.

We found that 3 earning groups have a different salary disparity between men and women (Figure 7). For the bottom group (<25k), there is no difference between genders, which we think cause these jobs to receive minimum wage salary, so between men and women there is no wage discrimination (Figure 4). However, as we get to the middle earning group (~50k), there is some difference between men and women. The linear regression line gives us a significant level of 0.000221 (Figure 5). All the black dots are on the right side, showing men always have a salary advantage. The same pattern repeats as we get to the top-earning group. For instance, within the same medical field, male physicians and surgeons earn, on average, 50k more, compared to women. We found that gender was a significant variable at 0.0348 level with a R^2 of 0.22 for this group (Figure 6). We also want to see this effect over the years. Figure 8 shows that the salary difference for the middle earning group more than doubles the low earning's one while the top-earning difference is more than fivefold for the low earning group. Over the year, across the group, there is a slight increase in the salary difference between men and women, except for the year 2015 when there are two financial crashes in the States.

For the STEM effect, initial data analysis shows us that, indicate that STEM majors do have a higher salary (Figure 9). Our regression model shows a significant difference between STEM and non-STEM majors at p=0.000 (Figure 10). In figure 11, the STEM majors do have a higher average median salary than non-STEM, with majors in physical science, healthcare, and engineering paid substantially higher than average.

For the model that combines both gender and STEM effects on earning, we see each of them strongly influences income (Figure 12). That the interaction between gender and STEM is not strong (Figure 13), and when being added, it doesn't change the R^2 value, meaning the interaction term between two variables doesn't play a significant role in explaining the salary differences. We think it makes sense because each contributes significantly to the response values that their interactions did not matter so much.

The result confirms our hypothesis that gender and STEM majors do significantly impact the salary range. In other words, there is a third variable that impacts both gender and STEM separately.

Appendix

Figure 1: EARNINGS AND GENDER RATIO
in year 2013, sized by total of men workers
in year 2013, sized by total of women workers
Earning($)
150000
100000
50000
0
25
50
75
100
Male-dominated (%)
Female-dominated (%)
0
500000
1000000
1500000
2000000
Major
Computer, Engineering, and Science
Education, Legal, Community Service, Arts, and Media
Healthcare Practitioners and Technical
Management, Business, and Financial
Natural Resources, Construction, and Maintenance
Production, Transportation, and Material Moving
Sales and Office
Service

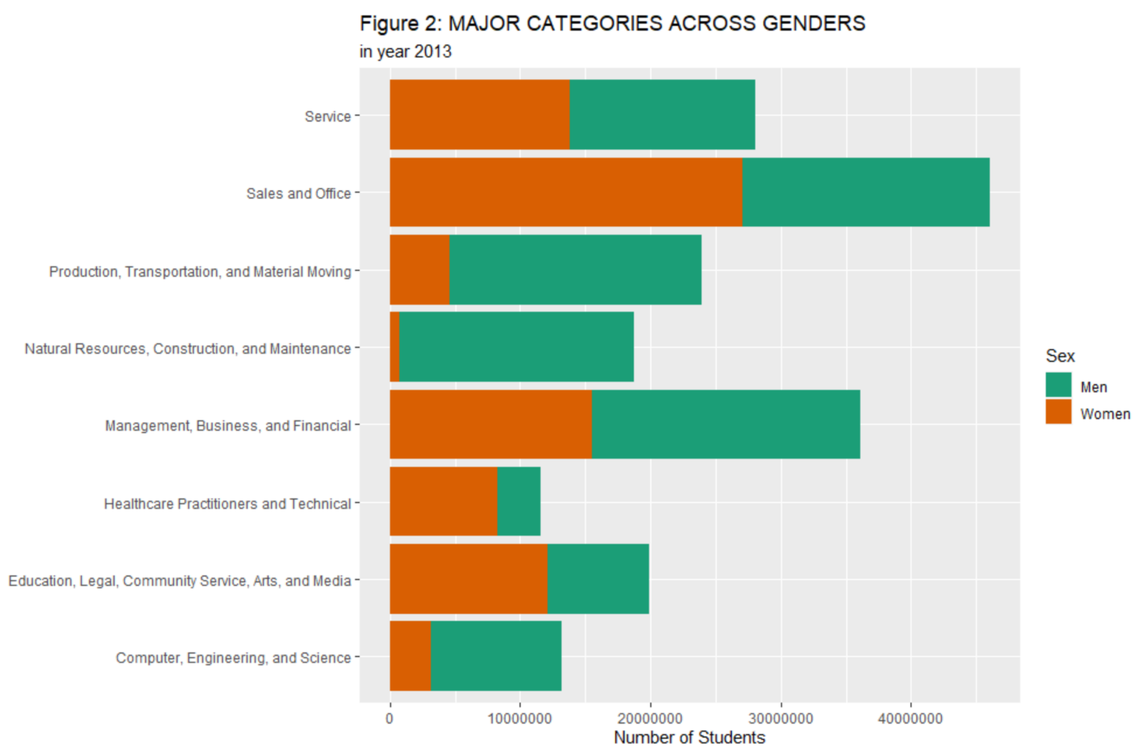
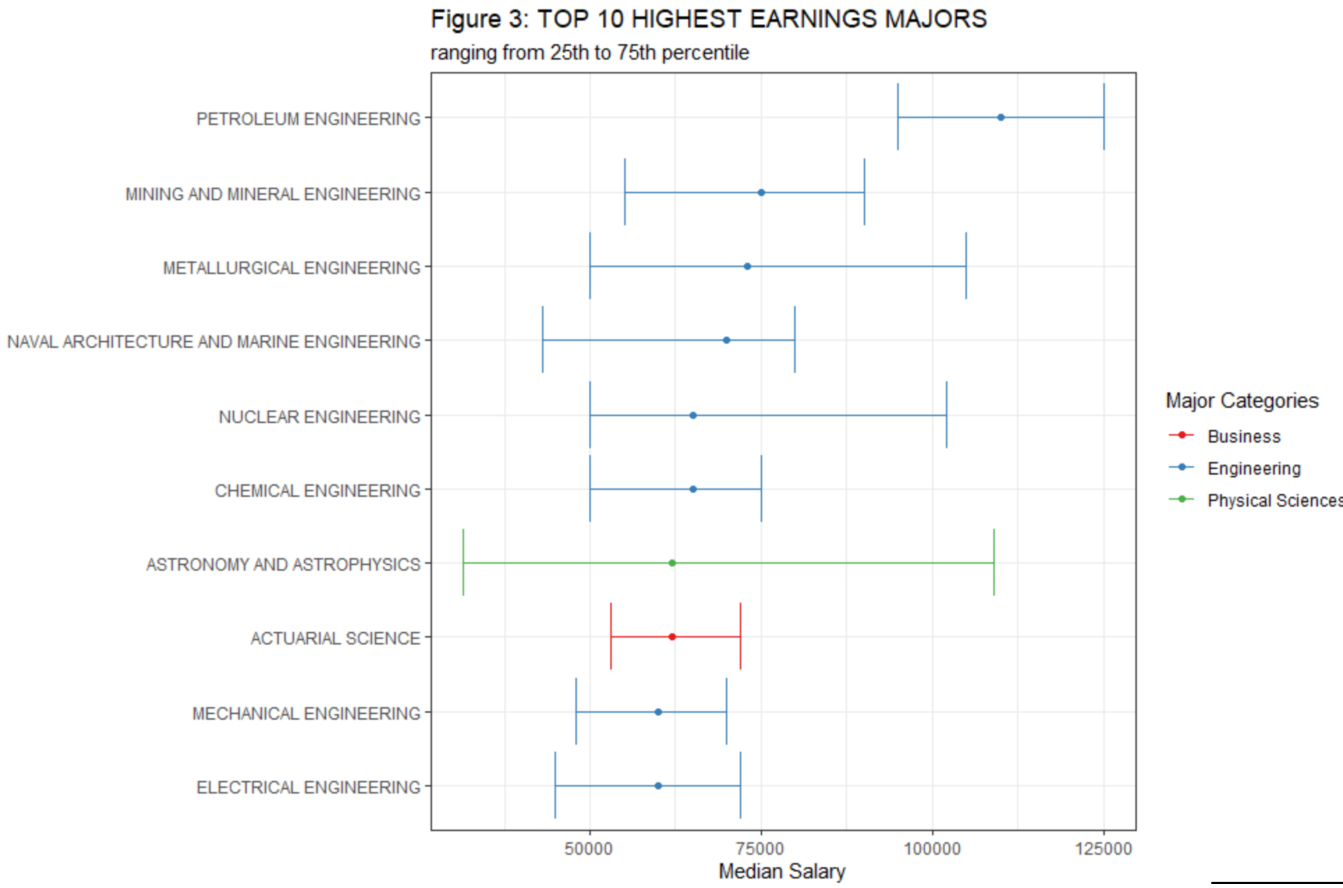


Figure 2: MAJOR CATEGORIES ACROSS GENDERS
in year 2013
Service
Sales and Office
Production, Transportation, and Material Moving
Natural Resources, Construction, and Maintenance
Management, Business, and Financial
Healthcare Practitioners and Technical
Education, Legal, Community Service, Arts, and Media
Computer, Engineering, and Science
0
10000000
20000000
30000000
40000000
Number of Students
Sex
Men
Women




```
Call:
lm(formula = Earning_both ~ Gender, data = earning_2013_lowcareer,
    subset = Gender %in% c("Men_earning", "Women_earning"))

Residuals:
    Min      1Q  Median      3Q     Max
-7851.0  -757.5   197.7  1197.1  3492.0

Coefficients:
                    Estimate Std. Error t value            Pr(>|t|)
(Intercept)          19998.0      779.5  25.656 0.00000000000000126
GenderWomen_earning  -1660.3     1102.3  -1.506               0.149

(Intercept)         ***
GenderWomen_earning
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2465 on 18 degrees of freedom
Multiple R-squared:  0.1119,    Adjusted R-squared:  0.06259
F-statistic: 2.269 on 1 and 18 DF,  p-value: 0.1494

>
```

Figure 4

```
Call:
lm(formula = Earning_both ~ Gender, data = earning_2013_middlecareer,
    subset = Gender %in% c("Men_earning", "Women_earning"))

Residuals:
    Min      1Q  Median      3Q     Max
-4331.1 -2734.0   556.1  1531.6  9203.9

Coefficients:
                    Estimate Std. Error t value             Pr(>|t|)
(Intercept)            47233       1129  41.826 < 0.0000000000000002 ***
GenderWomen_earning    -7657       1641  -4.667             0.000221 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3571 on 17 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.5616,    Adjusted R-squared:  0.5358
F-statistic: 21.78 on 1 and 17 DF,  p-value: 0.0002213
```

```
Call:
lm(formula = Earning_both ~ Gender, data = earning_2013_topcareer,
    subset = Gender %in% c("Men_earning", "Women_earning"))

Residuals:
   Min     1Q Median     3Q    Max
-47599 -13643  -7583  11835  67233

Coefficients:
                    Estimate Std. Error t value        Pr(>|t|)
(Intercept)           135300       7940  17.040 0.0000000000015 ***
GenderWomen_earning   -25636      11229  -2.283          0.0348 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 25110 on 18 degrees of freedom
Multiple R-squared:  0.2245,    Adjusted R-squared:  0.1815
F-statistic: 5.212 on 1 and 18 DF,  p-value: 0.03481
```

Figure 5 *Figure 6*

Figure 7: EARNING DIFFERENCE AMONGST EARNING GROUP ACROSS GENDERS

in 2013

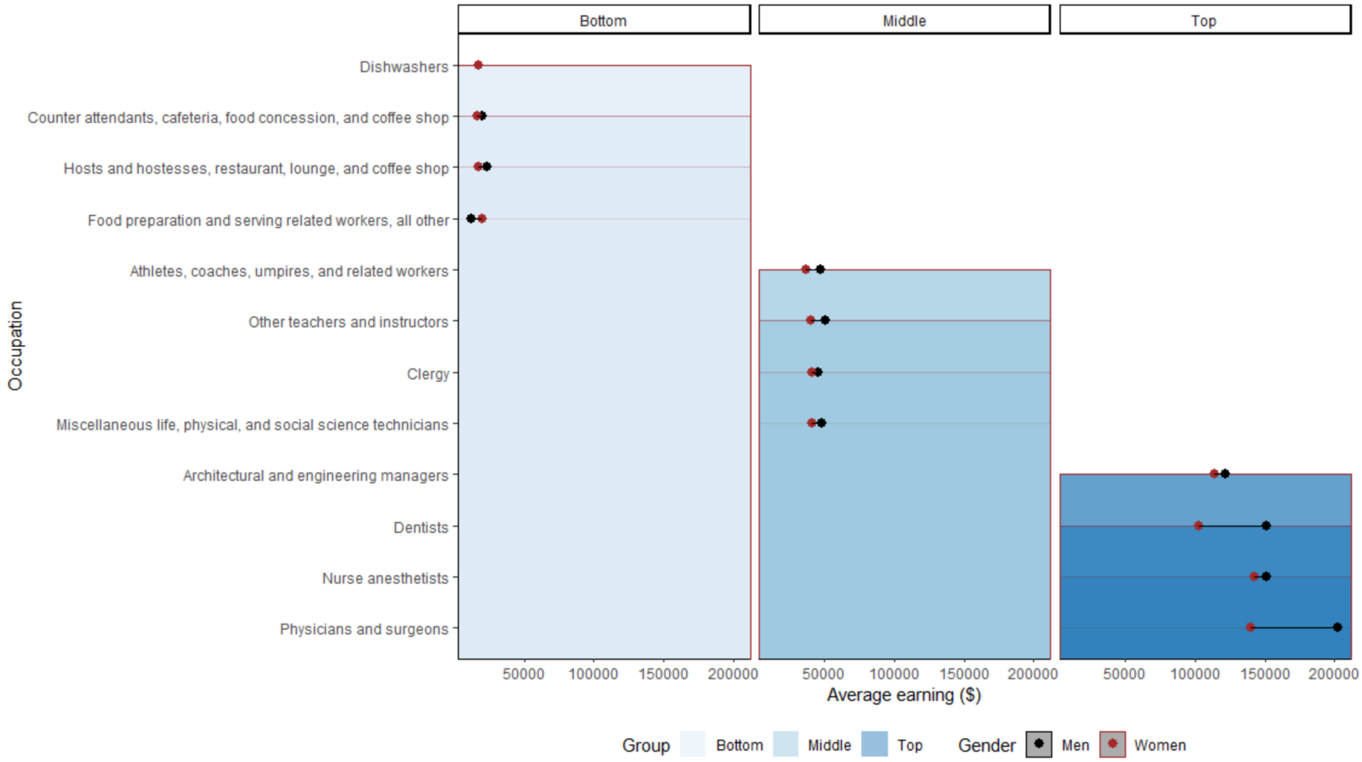
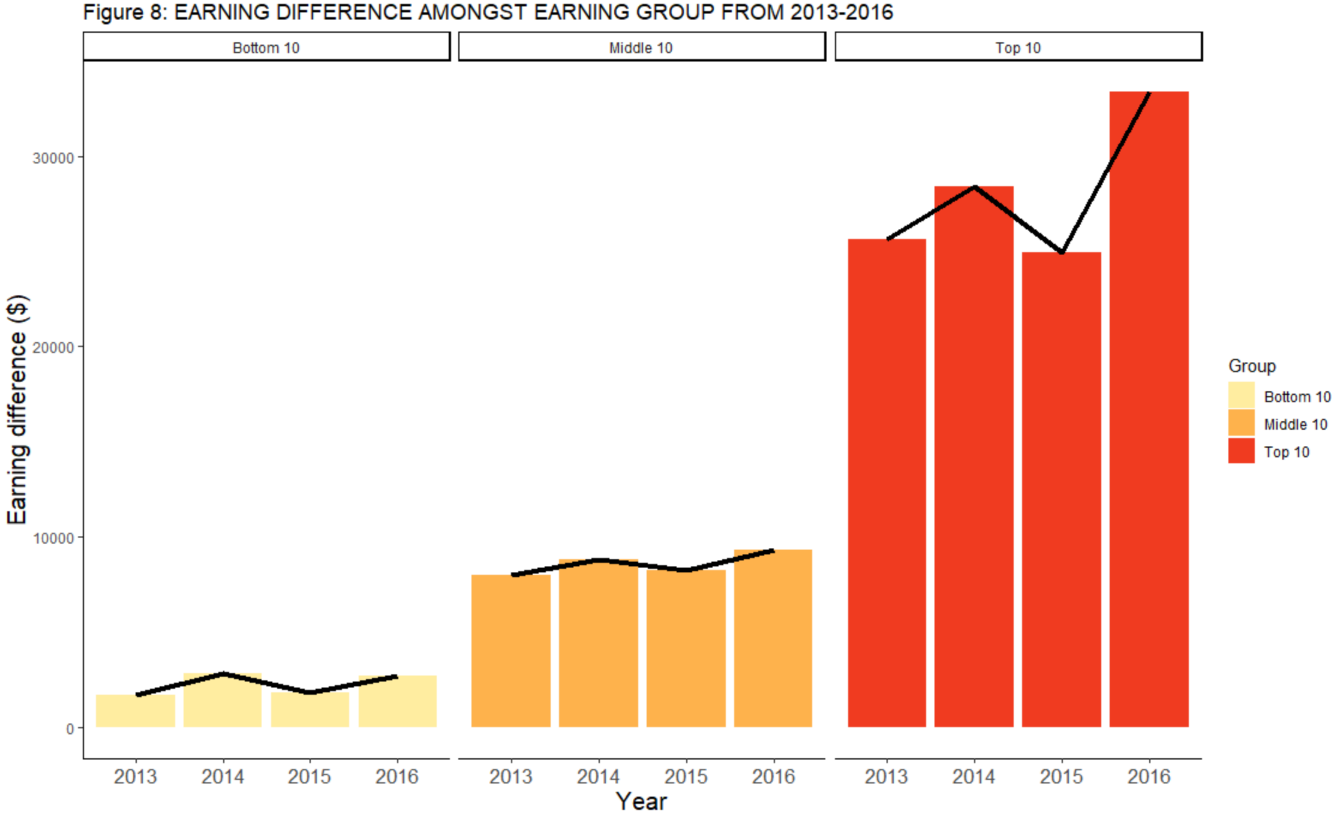
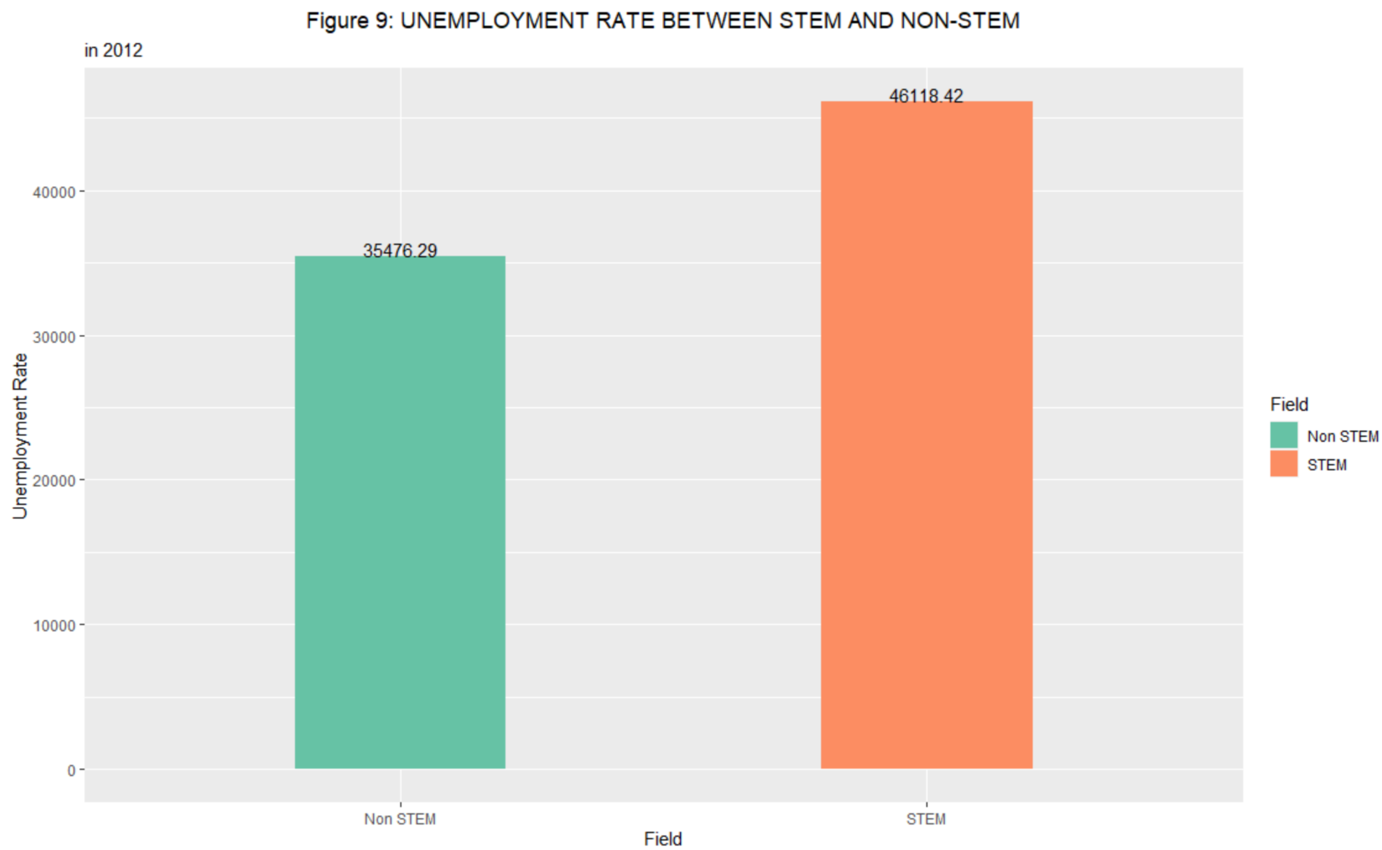


Figure 8: EARNING DIFFERENCE AMONGST EARNING GROUP FROM 2013-2016



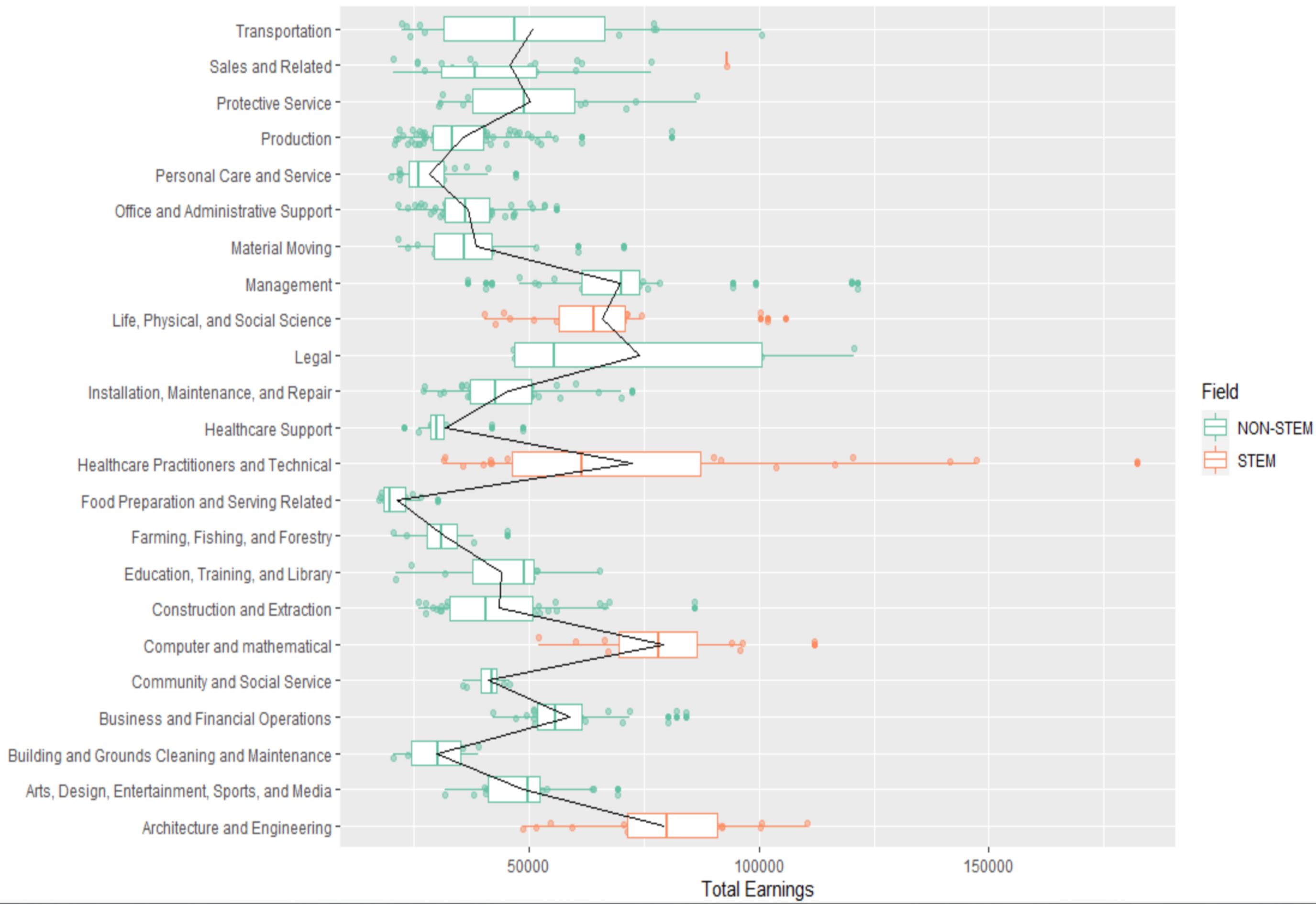



```
Call:
lm(formula = Earning ~ STEM, data = earning_2013)

Residuals:
   Min     1Q Median     3Q    Max
-42351 -12553  -2793   8265 108351

Coefficients:
            Estimate Std. Error t value            Pr(>|t|)
(Intercept)  43172.1      906.6   47.62 <0.0000000000000002 ***
STEMSTEM     30708.6     2159.5   14.22 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18800 on 520 degrees of freedom
Multiple R-squared:   0.28,	Adjusted R-squared:  0.2786
F-statistic: 202.2 on 1 and 520 DF,  p-value: < 0.00000000000000022
```

Figure 10

Figure 11: EARNINGS BETWEEN STEM AND NON-STEM

by major category, in 2013




```
Call:
lm(formula = Earning_both ~ Gender + STEM, data = earning_2013_gender)

Residuals:
   Min     1Q Median     3Q    Max 
-41399 -13460  -4312   9112 125494 

Coefficients:
                    Estimate Std. Error t value             Pr(>|t|)    
(Intercept)          46605.3      887.5  52.516 < 0.0000000000000002 ***
GenderWomen_earning  -8552.8     1202.4  -7.113      0.00000000000213 ***
STEMSTEM             30433.3     1570.9  19.373 < 0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19260 on 1024 degrees of freedom
Multiple R-squared:  0.2925,	Adjusted R-squared:  0.2911 
F-statistic: 211.7 on 2 and 1024 DF,  p-value: < 0.00000000000000022
```

Figure 12

```
Call:
lm(formula = Earning_both ~ Gender + STEM + Gender * STEM, data = earning_2013_gender)

Residuals:
   Min     1Q Median     3Q    Max 
-42615 -13489  -4191   9006 124278 

Coefficients:
                             Estimate Std. Error t value             Pr(>|t|)    
(Intercept)                     46348        929  49.887 < 0.0000000000000002 ***
GenderWomen_earning             -8028       1326  -6.052          0.000000002 ***
STEMSTEM                        31907       2223  14.353 < 0.0000000000000002 ***
GenderWomen_earning:STEMSTEM    -2944       3142  -0.937                0.349    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19270 on 1023 degrees of freedom
Multiple R-squared:  0.2931,	Adjusted R-squared:  0.291 
F-statistic: 141.4 on 3 and 1023 DF,  p-value: < 0.00000000000000022
```

Figure 13

Reference

Carnevale, A.P.,Strohl, J, & Melton, M. (2013). What's It Worth?: The Economic Value of College Majors. Georgetown University center for Education and Workforce. http://hdl.handle.net/10822/559309

Dowell, E.K.P. (2022). Women consistently earn less than men. https://www.census.gov/library/stories/2022/01/gender-pay-gap-widens-as-women-age.html

Handelman, J. & Smith, G. (2016). STEM for all. https://obamawhitehouse.archives.gov/blog/2016/02/11/stem-all doi:https://doi.org/10.1016/S0272-7757(01)00054-1

Data: https://github.com/fivethirtyeight/data/tree/master/college-majors
https://github.com/rfordatascience/tidytuesday/blob/master/data/2019/2019-03-05/jobs_gender.csv