

# Assignment\_Answers

Sachin Sharma

10/7/2021

## Importing Libraries

```
library(tidyr)
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5      v dplyr 1.0.7
## v tibble 3.1.4       v stringr 1.4.0
## v readr 2.0.1        v forcats 0.5.1
## v purrr 0.3.4

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library(dplyr)
library(ggplot2)
library(corrgram)
library(corrplot)

## corrplot 0.90 loaded

library(tinytex)

df <- read.csv("college.csv")
```

## Checking the data

```
head(df)
```

##	X	rank	major_code	major	major_category			
## 1	1	1	2419	Petroleum Engineering	Engineering			
## 2	2	2	2416	Mining And Mineral Engineering	Engineering			
## 3	3	3	2415	Metallurgical Engineering	Engineering			
## 4	4	4	2417	Naval Architecture And Marine Engineering	Engineering			
## 5	5	5	2405	Chemical Engineering	Engineering			
## 6	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
```

```
view(df)
```

To handle this data to answer our questions, we need to arrange data properly by using the following codes :

```
df_1 <- df[order(df$perc_low_wage_jobs),]
```

```
head(df_1)
```

```
##      X rank major_code      major
## 21  21  21      2102      Computer Science
## 26  26  26      2406      Civil Engineering
## 74  74  74      3801      Military Technologies
## 111 111 111      5002 Atmospheric Sciences And Meteorology
## 120 120 120      2305      Mathematics Teacher Education
## 53  53  53      4005      Mathematics And Computer Science
##      major_category total sample_size perc_women p25th
## 21      Computers & Mathematics 128319      1196 0.5155433 36000
## 26      Engineering 53153      565 0.1789819 30000
## 74 Industrial Arts & Consumer Services 124      4 0.3422288 25000
## 111      Physical Sciences 4043      32 0.7920953 48000
## 120      Education 14237      123 0.1073132 23050
## 53      Computers & Mathematics 609      7 0.6516599 24000
##      median p75th perc_men perc_employed perc_employed_fulltime
## 21  45000 50000 0.4844567      0.8486273      0.8019163
## 26  33000 45000 0.8210181      0.8192260      0.7910734
## 74  35600 40200 0.6577712      0.7295205      0.6711353
## 111 60000 70000 0.2079047      0.9561650      0.7443252
## 120 34000 42000 0.8926868      0.5535322      0.8057198
```

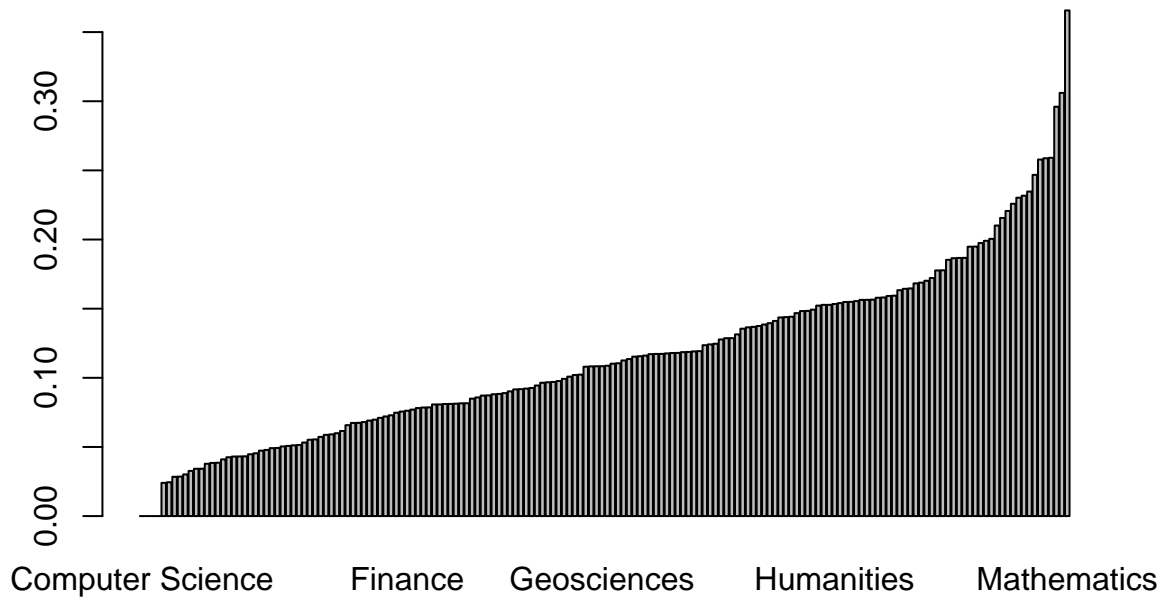
```
## 53    33400 45000 0.3483401    0.7684569    0.8367065
##      perc_employed_parttime perc_employed_fulltime_yearround perc_unemployed
## 21              0.3536763              0.8104983    0.15137274
## 26              0.2919774              0.7097022    0.18077399
## 74              0.4657461              0.7799241    0.27047948
## 111             0.4543145              0.7283242    0.04383502
## 120             0.3498961              0.6638572    0.44646782
## 53              0.2851886              0.7189205    0.23154311
##      perc_college_jobs perc_non_college_jobs perc_low_wage_jobs
## 21              0.6972603              0.3027397    0.00000000
## 26              0.6978892              0.1345646    0.00000000
## 74              0.7037037              0.2716049    0.00000000
## 111             0.5791191              0.2349103    0.00000000
## 120             0.3744186              0.4852713    0.02403101
## 53              0.7940732              0.1182340    0.02449350
```

The above chunk of code has arranged the last column which is `perc_low_wage_jobs` in ascending order, which will help us to answer our question of what is the correlation between major subject and low wage jobs.

Now let us plot a barplot to see the relation with the help of our data frame `df_1`

```
barplot(df_1$perc_low_wage_jobs, names.arg = df_1$major, main = "Low Wage Jobs Vs. Major Subject" )
```

## Low Wage Jobs Vs. Major Subject



The above graph is showing that there is strong correlation between major subject and low wage income.

Question: Based on your analysis, would you conclude that there is a significant association between college major category and income?

Answer : With the help of the graph , we can say that there is significant association between college major category and income.

Question : Please type a few sentences describing your results.

As we can see with the help of graph that bars are increasing as there is change in the major subjects viz. Computer Science, Mathematics, .....and so on.. We can see that minimum wage salary has been increasing as there is change in major subjects. This graph show us that there is very strong association between College major category and income.