Final Report on Campus Recruitment

Academic and Employ ability Factors Influencing Placement

MATH 4330

December 2021

Contents

#Introduction

I am finding this data set to discover this interesting causal question: Which factor influenced a candidate in getting placed or not in campus recruitment. This data set consists of Placement data of students in a ordinary campus. It includes secondary and higher secondary school percentage and specialization. It also includes degree specialization, type and Work experience and salary offers to the placed students There are total 215 students being survery in this data set.

##Independent Variable

Gender has two categories here, Male='M',Female='F'.
ssc_p means Secondary Education percentage at 10th Grade, it is a continuous independent variable. ssc_b means Board of Education, it is a categorical independent variable with category as Central or Others.

hsc_p measn Higher Secondary Education percentage at 12th Grade, it is a continuous independent variable. hsc_b means Board of Education. it is a categorical independent variable with category as Central or Others. hsc_s means Specialization in Higher Secondary Education. It has three categories, commence, science and art

degree_p means Degree Percentage.it is a continuous independent variable. degree_t means Under Graduation(Degree type)/Field of degree education. It has three categories.Commence & Management, Science & Technology,others.

workex means Work Experience. It has categories of Yes and No. etest_p means Employ ability Test Percentage. it is a continuous independent variable. specialisation means what major they study. it has 2 categories, market&finance, market&human resource. mba_p means MBA Percentage. it is a continuous independent variable. salary is what the student earn if they find a placement

#Dependent Variable

status is the dependent variable. it have two categories, placed or not placed

```
Placement_Data_Full_Class <- read.csv("C:/Users/yorkuniversity/Desktop/archive/Placement_Data_Full_Clas
library(spida2)
library(lattice)
```

```
## Warning: package 'lattice' was built under R version 4.0.5
```

```r
library(latticeExtra)
```

```
## Warning: package 'latticeExtra' was built under R version 4.0.5
```

```r
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5

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

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1

## Warning: package 'ggplot2' was built under R version 4.0.5

## Warning: package 'tibble' was built under R version 4.0.5

## Warning: package 'tidyr' was built under R version 4.0.5

## Warning: package 'readr' was built under R version 4.0.5

## Warning: package 'dplyr' was built under R version 4.0.5

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x readr::cols()    masks spida2::cols()
## x dplyr::filter()  masks stats::filter()
## x ggplot2::labs()  masks spida2::labs()
## x dplyr::lag()     masks stats::lag()
## x ggplot2::layer() masks latticeExtra::layer()
## x purrr::map()     masks spida2::map()
```

```r
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 4.0.5

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```

```r
library(plyr)
```

```
## ------------------------------------------------------------------------------
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)

## ------------------------------------------------------------------------------

##
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

## The following object is masked from 'package:purrr':
##
##     compact

## The following object is masked from 'package:spida2':
##
##     here
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.0.5

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

## The following object is masked from 'package:purrr':
##
##     some
```

```
library(spida2)
library(pROC)
```

```
## Warning: package 'pROC' was built under R version 4.0.5

## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```r
library(broom)
```

```
## Warning: package 'broom' was built under R version 4.0.5
```

```r
library(MASS)
```

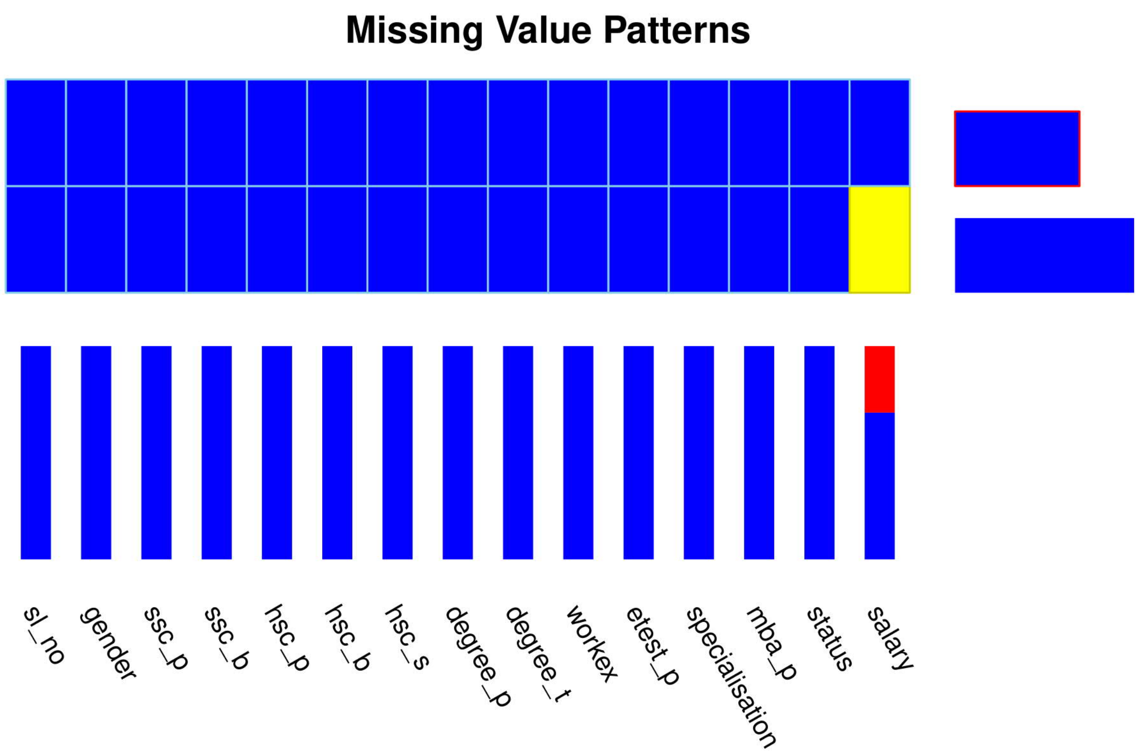
```
##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

## The following object is masked from 'package:spida2':
##
##     Null
```

```r
library(plotly)
```

```
## Warning: package 'plotly' was built under R version 4.0.5

##
## Attaching package: 'plotly'

## The following object is masked from 'package:MASS':
##
##     select

## The following objects are masked from 'package:plyr':
##
##     arrange, mutate, rename, summarise

## The following object is masked from 'package:ggplot2':
##
##     last_plot

## The following object is masked from 'package:stats':
##
##     filter

## The following object is masked from 'package:graphics':
##
##     layout
```

#Closer look of the data

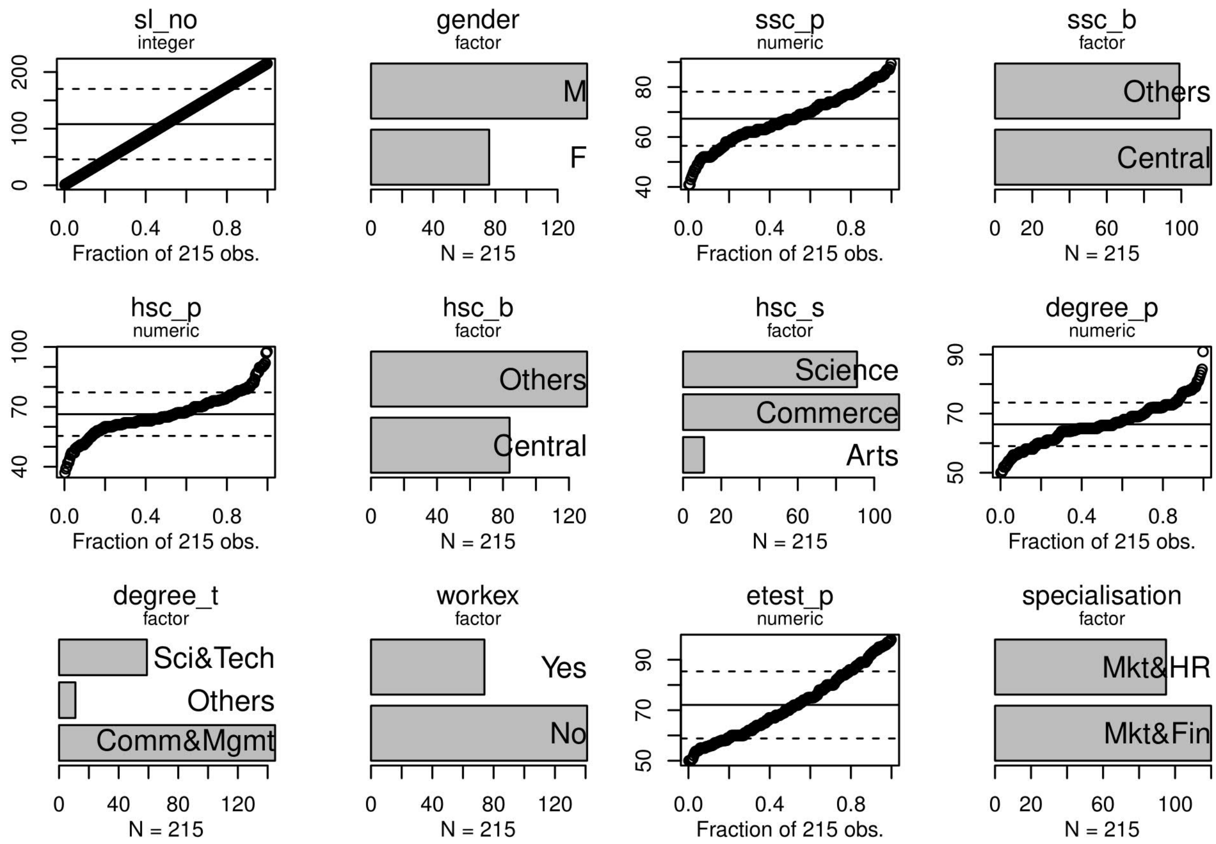
```r
tablemissing(Placement_Data_Full_Class)
```

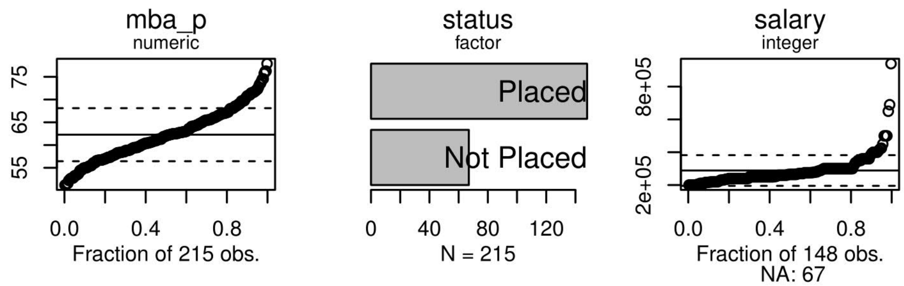



```
##       sl_no gender ssc_p ssc_b hsc_p hsc_b hsc_s degree_p degree_t workex
## 1         1      1     1     1     1     1     1        1        1      1
## 2         1      1     1     1     1     1     1        1        1      1
## Total     0      0     0     0     0     0     0        0        0      0
##       etest_p specialisation mba_p status salary Total
## 1           1              1     1      1      1   148
## 2           1              1     1      1      0    67
## Total       0              0     0      0     67   215
```

using this function to check if we have missing value, the red color in salary shows we have missing value for salary but it makes sense because those miss salary are the people who do not find the placement, thus, they have missing value for salary

```
xqplot(Placement_Data_Full_Class)
```


sl_no
integer
Fraction of 215 obs.
gender
factor
M
F
N = 215
ssc_p
numeric
Fraction of 215 obs.
ssc_b
factor
Others
Central
N = 215
hsc_p
numeric
Fraction of 215 obs.
hsc_b
factor
Others
Central
N = 215
hsc_s
factor
Science
Commerce
Arts
N = 215
degree_p
numeric
Fraction of 215 obs.
degree_t
factor
Sci&Tech
Others
Comm&Mgmt
N = 215
workex
factor
Yes
No
N = 215
etest_p
numeric
Fraction of 215 obs.
specialisation
factor
Mkt&HR
Mkt&Fin
N = 215

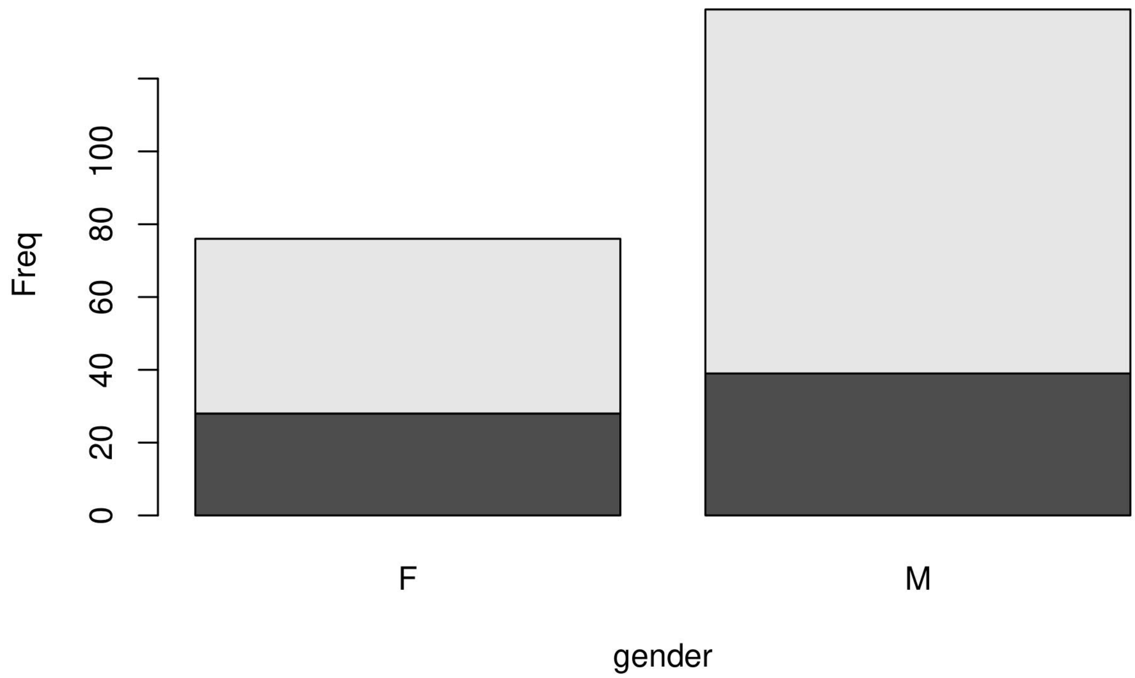


xqplot function is also useful for giving me a big picture of what is going on we see all the continuous line and nice curves in the percentage scores it shows the data set are nicely distributed, there is no missing data. it also shows the different categories in each categorical data.

```
ddd<-as.data.frame(tab__(~status+gender,Placement_Data_Full_Class))
ddd
```

```
##       status gender Freq
## 1 Not Placed      F   28
## 2     Placed      F   48
## 3 Not Placed      M   39
## 4     Placed      M  100
```

```
barplot(Freq~status+gender,ddd)
```




```r
#barplot(status~gender,Placement_Data_Full_Class,horiz = FALSE, xlab = NULL, ylab = NULL)
```

this indicates to us the distribution of table to gender and see how many with female or male get placement overall

#Trial out analysis

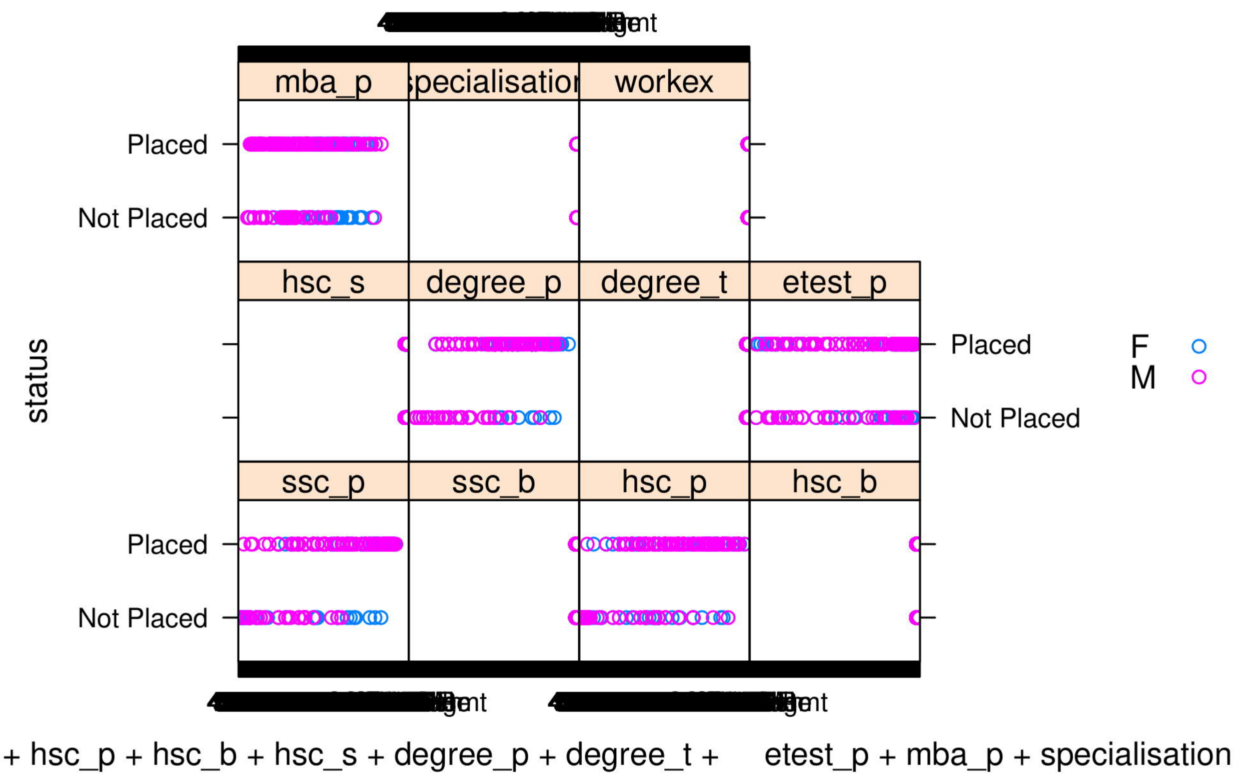
```r
head(Placement_Data_Full_Class)
```

```
##   sl_no gender ssc_p   ssc_b hsc_p   hsc_b    hsc_s degree_p  degree_t workex
## 1     1      M 67.00  Others 91.00  Others Commerce    58.00  Sci&Tech     No
## 2     2      M 79.33 Central 78.33  Others  Science    77.48  Sci&Tech    Yes
## 3     3      M 65.00 Central 68.00 Central     Arts    64.00 Comm&Mgmt     No
## 4     4      M 56.00 Central 52.00 Central  Science    52.00  Sci&Tech     No
## 5     5      M 85.80 Central 73.60 Central Commerce    73.30 Comm&Mgmt     No
## 6     6      M 55.00  Others 49.80  Others  Science    67.25  Sci&Tech    Yes
##   etest_p specialisation mba_p     status salary
## 1    55.0         Mkt&HR 58.80     Placed 270000
## 2    86.5        Mkt&Fin 66.28     Placed 200000
## 3    75.0        Mkt&Fin 57.80     Placed 250000
## 4    66.0         Mkt&HR 59.43 Not Placed     NA
## 5    96.8        Mkt&Fin 55.50     Placed 425000
## 6    55.0        Mkt&Fin 51.58 Not Placed     NA
```

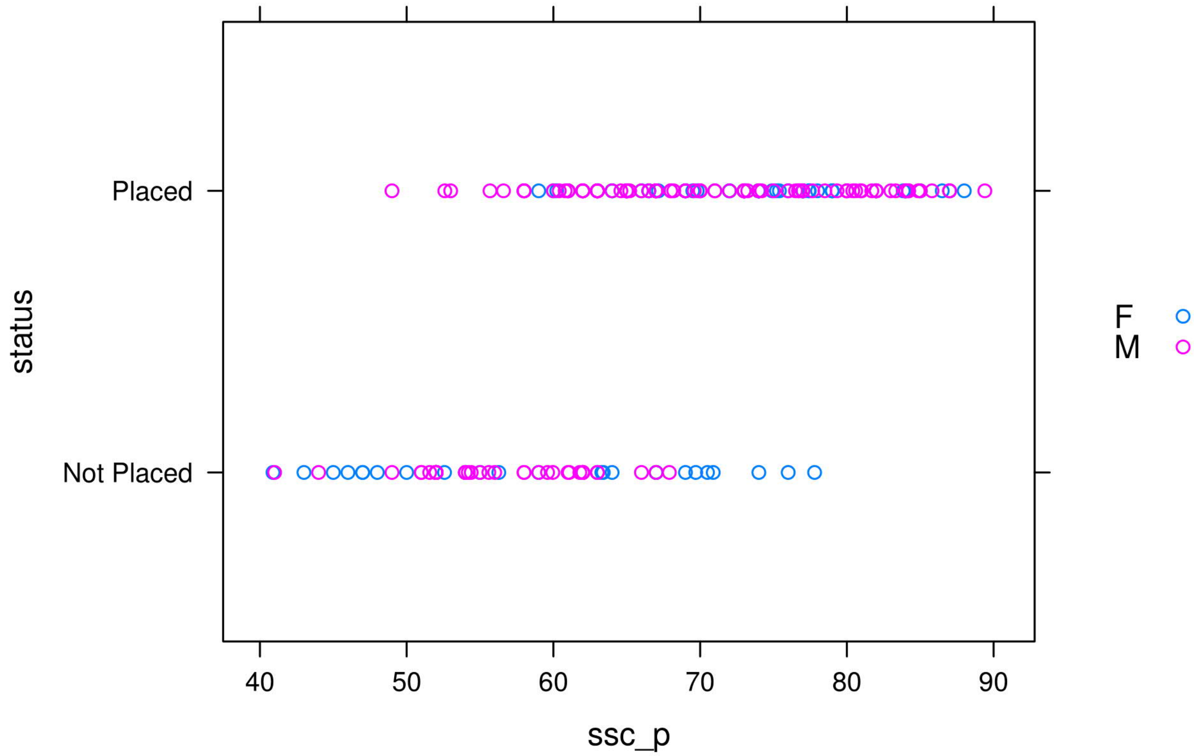
I think the xyplot function can give me some ideas on what factor is important to have and it can give some indications that determined on how you group the information, so when you group the information differently, we will be able to see the difference in between them.

the following shows group by gender, but we can see it is hard to see whether there is discrimination or not.

```
xyplot(status ~ ssc_p +ssc_b + hsc_p + hsc_b+ hsc_s + degree_p + degree_t+etest_p+
                mba_p+ specialisation+workex, Placement_Data_Full_Class, groups = gender,
       auto.key= list(space='right'))
```

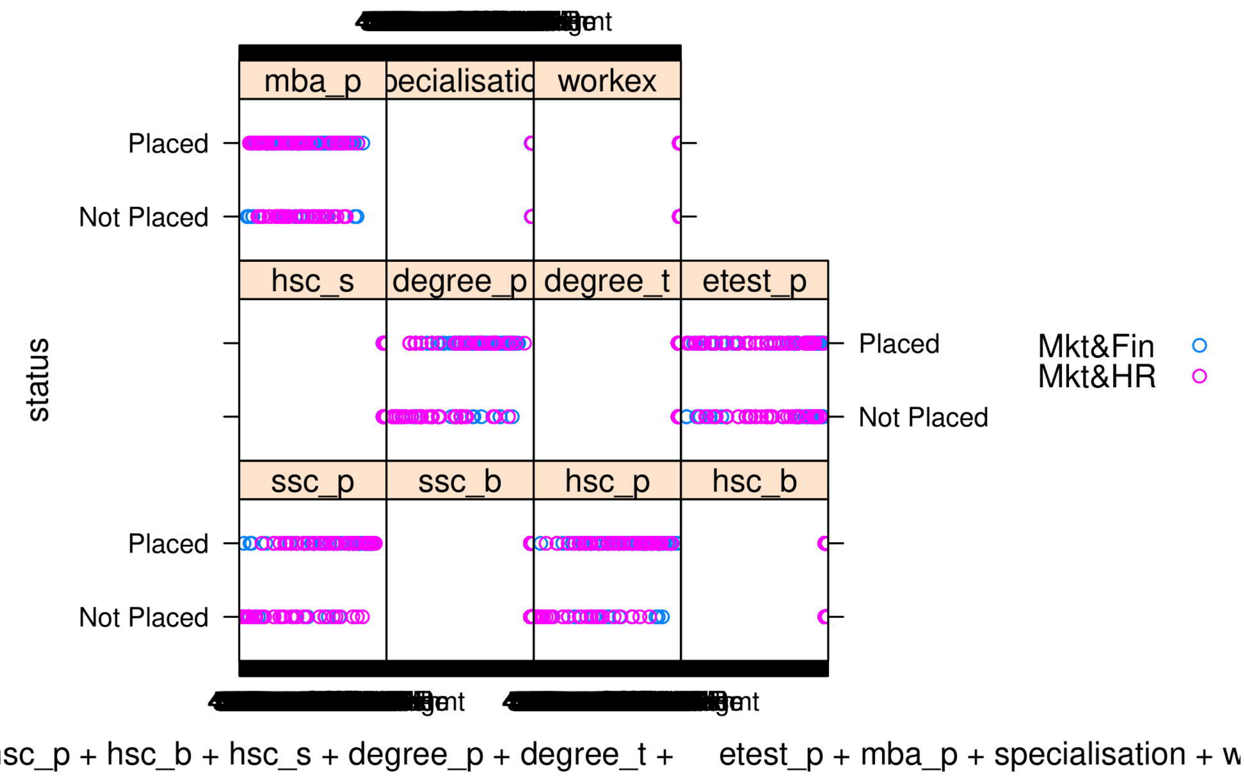



```
xyplot(status ~ ssc_p, Placement_Data_Full_Class, groups = gender,
       auto.key= list(space='right'))
```

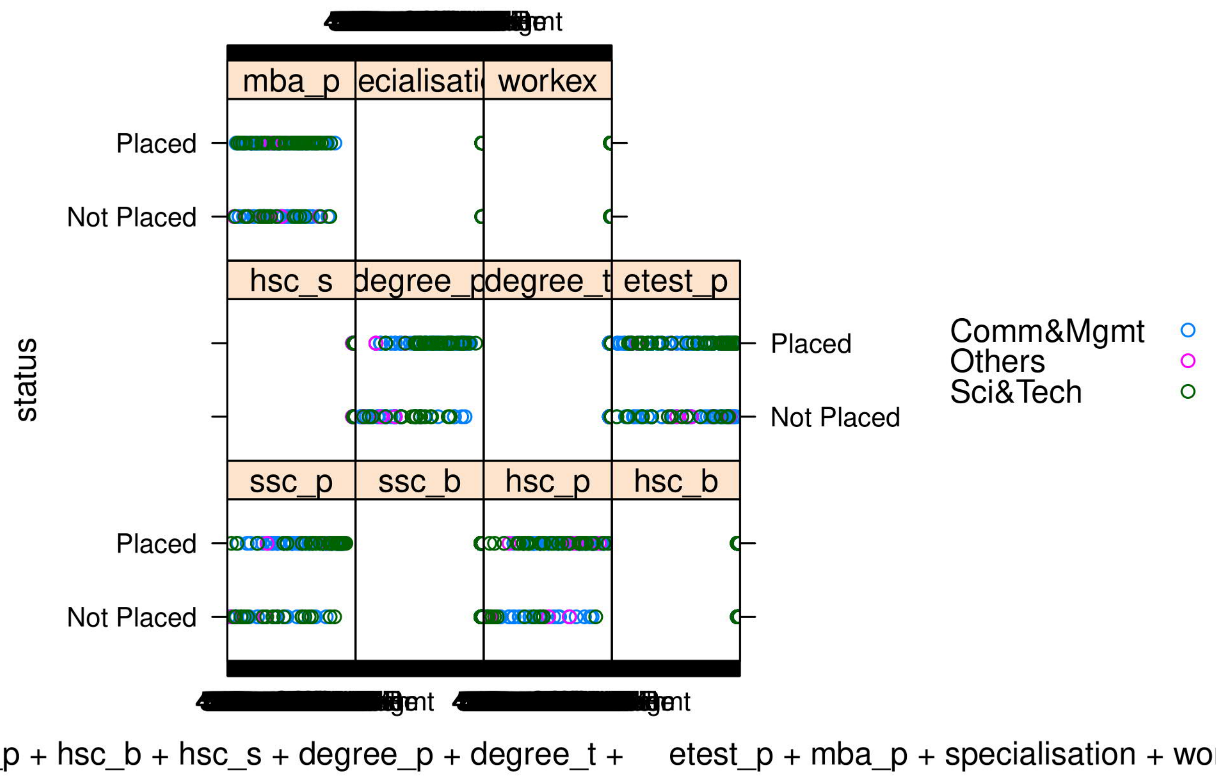


I have also try with other grouping methods. Specialisation and major choice and level of school being a good school or ok school does not affect much when you observe the number of students get placement or not

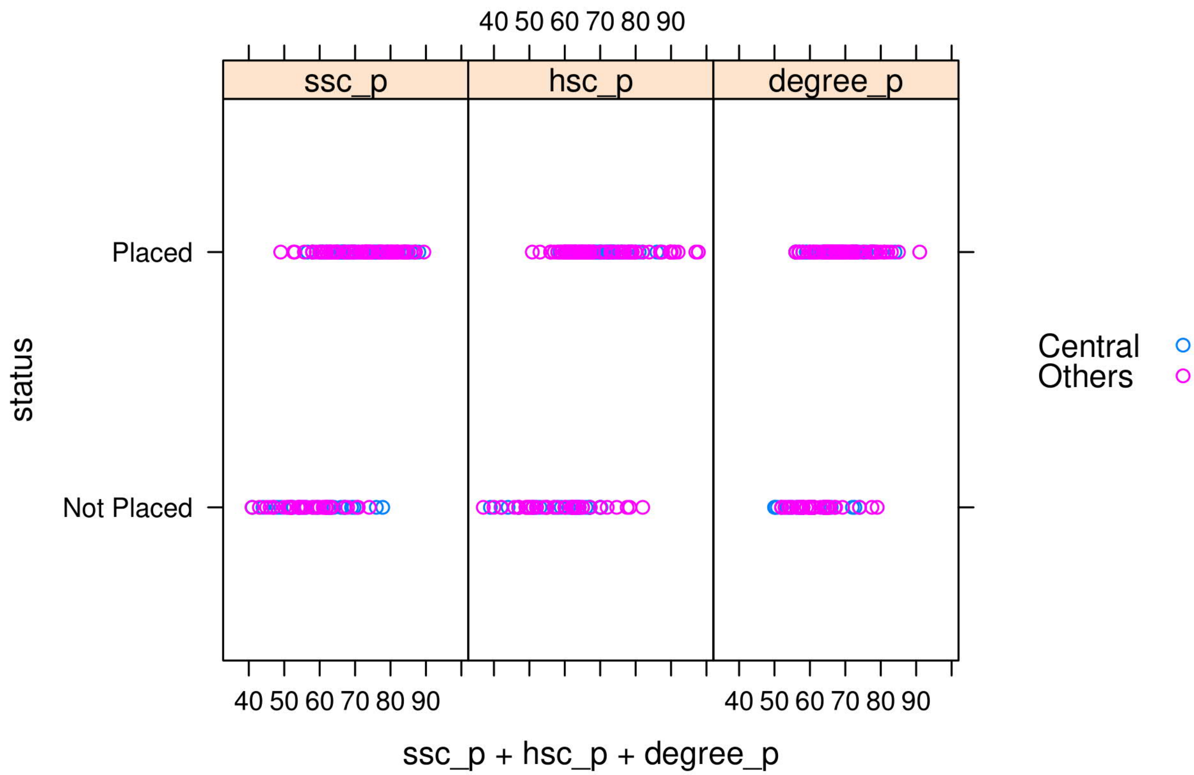
```
xyplot(status ~ ssc_p +ssc_b + hsc_p + hsc_b+ hsc_s + degree_p + degree_t+etest_p+
                  mba_p+ specialisation+workex, Placement_Data_Full_Class, groups = specialisation,
        auto.key= list(space='right'))
```




```
xplot(status ~ ssc_p +ssc_b + hsc_p + hsc_b+ hsc_s + degree_p + degree_t+etest_p+
                mba_p+ specialisation+ workex, Placement_Data_Full_Class, groups = degree_t,
       auto.key= list(space='right'))
```

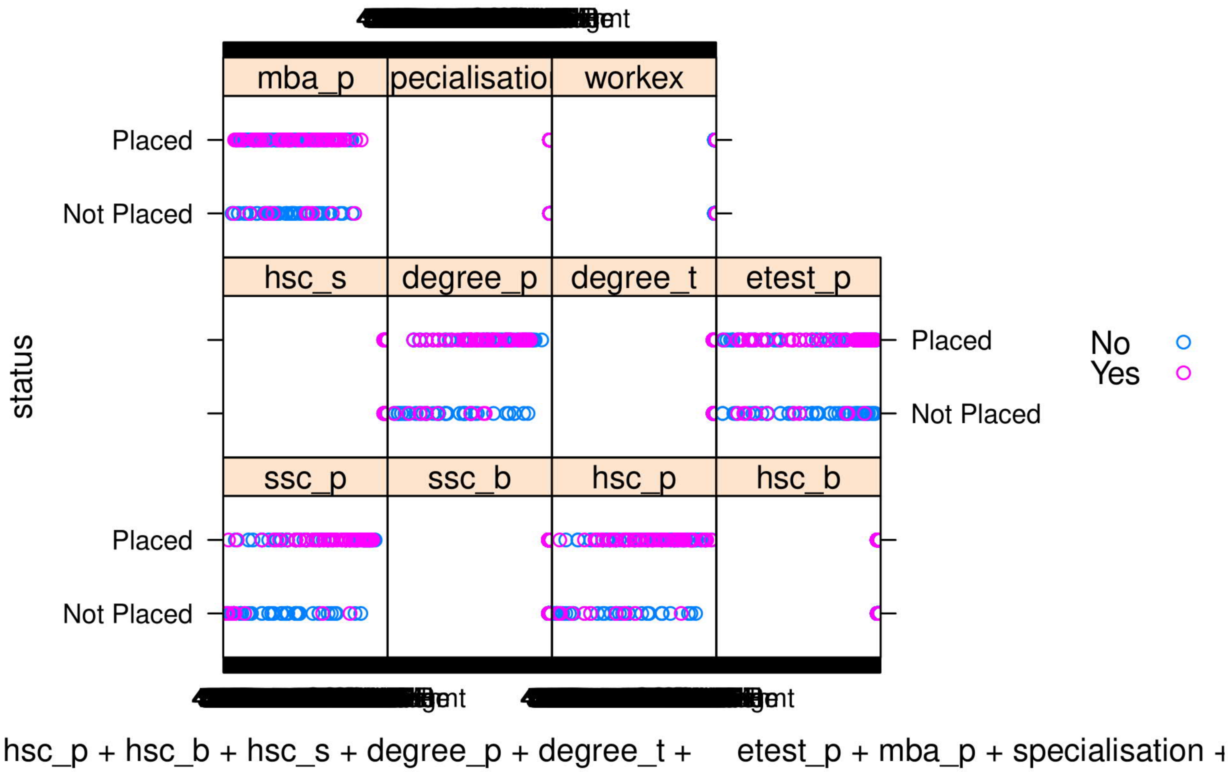



```
xyplot(status ~ ssc_p +hsc_p+degree_p, Placement_Data_Full_Class, groups =hsc_b ,
       auto.key= list(space='right'))
```



but from there you can see work experience do place a role a bit more significant compared with others, we can see the pink color a bit more clear that it distributed in the range of finding a job compared to those who do not have work experience do not have a placement

```
xyplot(status ~ ssc_p +ssc_b + hsc_p + hsc_b+ hsc_s + degree_p + degree_t+etest_p+
                 mba_p+ specialisation+ workex, Placement_Data_Full_Class, groups = workex,
        auto.key= list(space='right'))
```

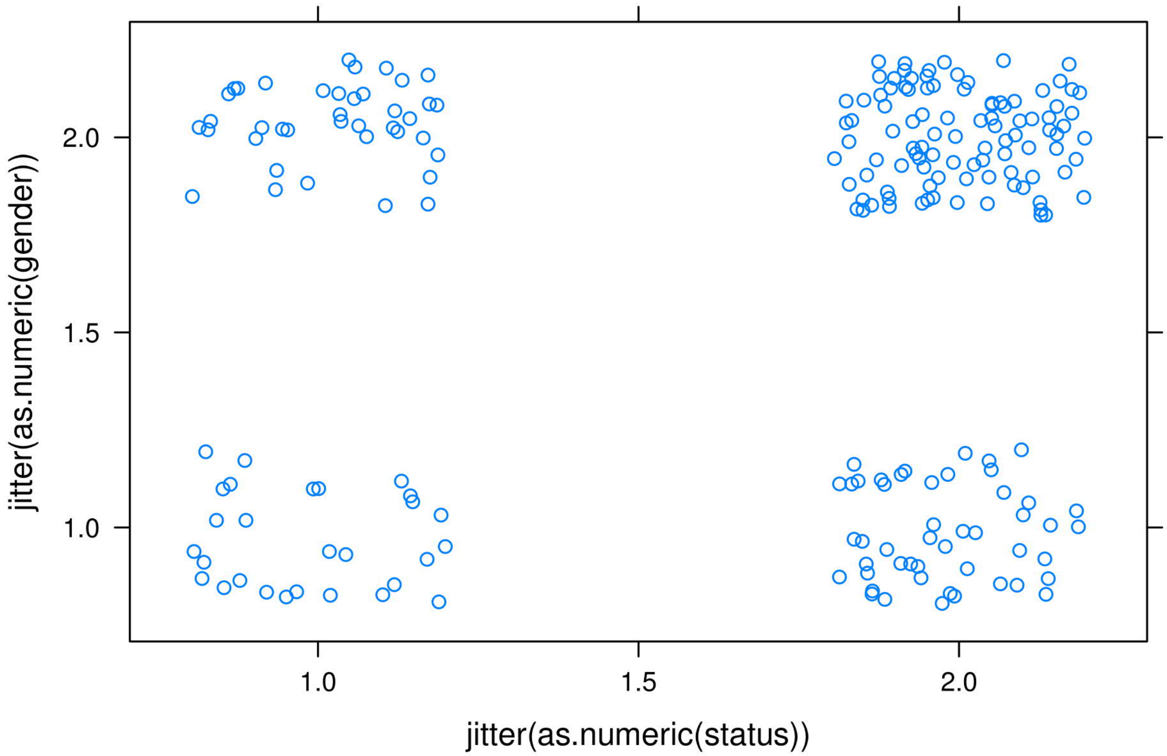


Thus, after try out and see how the data trend looks like,I discovers that is important for which group we are grouping for. From our test, we can see people with work experience tends to get placement. Male has a slightly advantage to get placement more than female, but not it does not show as very significant school level does not affect much too because we see people from both central and others get placement.

Data presentation

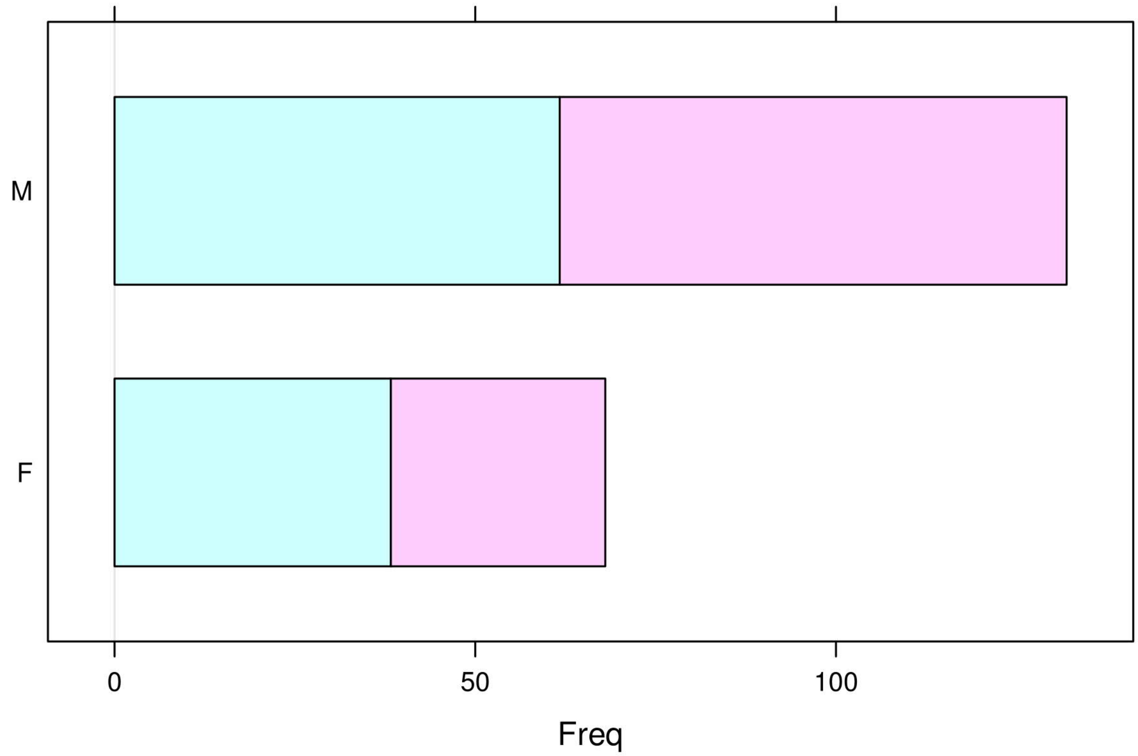
The following ones also shows some information for this data set that may help us see what are important factors to include in the model.it gives us a different data presentation issue

```
status<-as.factor(Placement_Data_Full_Class$status)
xyplot(jitter(as.numeric(gender))~jitter(as.numeric(status)), Placement_Data_Full_Class)
```



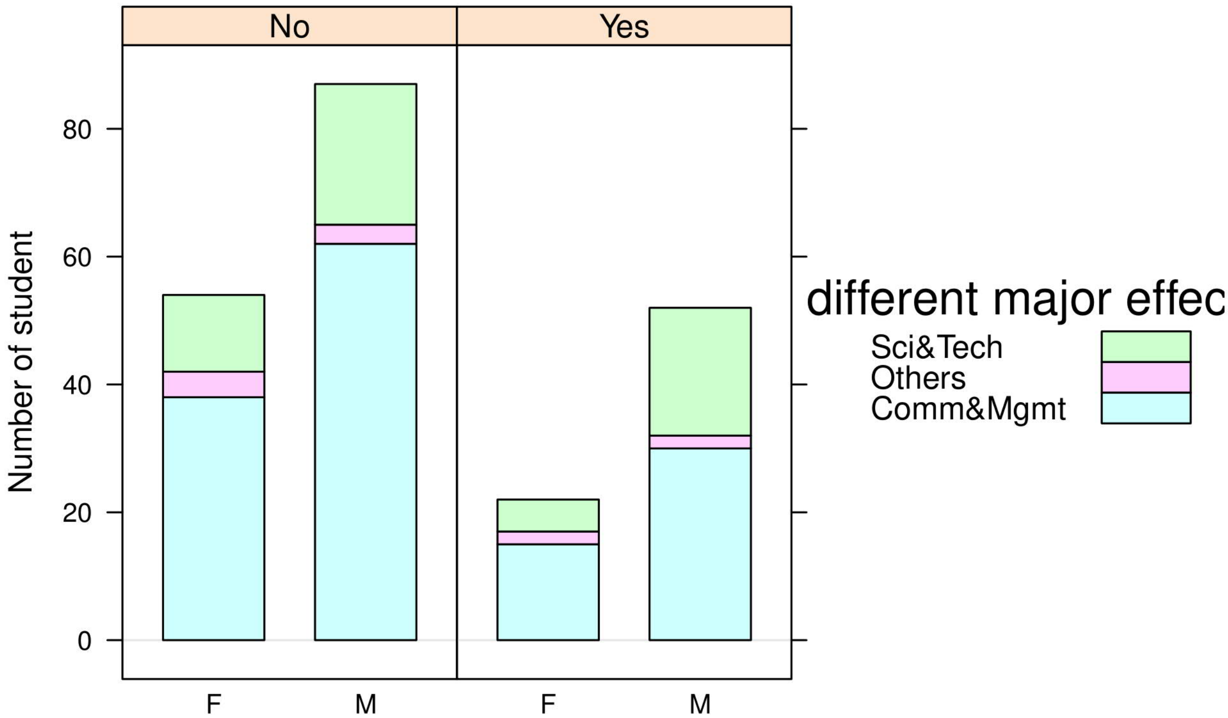
this shows how gender distributed over getting placement or not

```
#tab__(status~gender, Placement_Data_Full_Class)

tab__(Placement_Data_Full_Class, ~ gender + workex, pct = 2, test = TRUE)%>%
        barchart
```



this barchart shows among different gender, how many get placement

```
tab__(~gender + workex + degree_t, Placement_Data_Full_Class) %>%
        barchart(
                horizontal = FALSE,
                ylab = 'Number of student',
                auto.key = list(
                        space = 'right',
                        reverse.rows = T,
                        title = 'different major effect '
                )
        )
```



this shows among different gender, for students who get placement what do they study in major in university, we can clearly see commence and management people tends to get placement for both gender.

Over this rough trails, we know work experience, major choice, and gender play some effects in whether people find placement or not. let's do deeper investigation now.

#Analysis on Coefficient

```
fit<- glm(status~ gender+ssc_p +ssc_b + hsc_p + hsc_b+ hsc_s + degree_p + degree_t+etest_p+
                  mba_p+ specialisation+workex, Placement_Data_Full_Class,family=binomial)

summary(fit)
```

```
##
## Call:
## glm(formula = status ~ gender + ssc_p + ssc_b + hsc_p + hsc_b +
##     hsc_s + degree_p + degree_t + etest_p + mba_p + specialisation +
##     workex, family = binomial, data = Placement_Data_Full_Class)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.30074  -0.14447   0.07164   0.31692   2.32685
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -18.37171    5.32356  -3.451 0.000558 ***
## genderM                1.19433    0.68598   1.741 0.081673 .
```

```
## ssc_p                  0.22891    0.04682   4.889 1.01e-06 ***
## ssc_bOthers            0.22767    0.71685   0.318 0.750787
## hsc_p                  0.10721    0.03778   2.838 0.004541 **
## hsc_bOthers            0.33074    0.73509   0.450 0.652757
## hsc_sCommerce         -1.49787    1.36117  -1.100 0.271143
## hsc_sScience          -0.91121    1.45714  -0.625 0.531746
## degree_p               0.18577    0.05558   3.343 0.000830 ***
## degree_tOthers        -1.11791    1.54778  -0.722 0.470132
## degree_tSci&Tech      -1.72576    0.79258  -2.177 0.029452 *
## etest_p               -0.01416    0.02266  -0.625 0.532060
## mba_p                 -0.21413    0.05852  -3.659 0.000253 ***
## specialisationMkt&HR  -0.26381    0.55610  -0.474 0.635217
## workexYes              2.08385    0.70839   2.942 0.003264 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 266.771  on 214  degrees of freedom
## Residual deviance:  99.677  on 200  degrees of freedom
## AIC: 129.68
##
## Number of Fisher Scoring iterations: 7
```

```r
fit2<-stepAIC(fit)
```

```
## Start:  AIC=129.68
## status ~ gender + ssc_p + ssc_b + hsc_p + hsc_b + hsc_s + degree_p +
##     degree_t + etest_p + mba_p + specialisation + workex
##
##                  Df Deviance    AIC
## - hsc_s           2  101.327 127.33
## - ssc_b           1   99.778 127.78
## - hsc_b           1   99.879 127.88
## - specialisation  1   99.901 127.90
## - etest_p         1  100.069 128.07
## <none>               99.677 129.68
## - gender          1  102.855 130.85
## - degree_t        2  105.236 131.24
## - hsc_p           1  109.778 137.78
## - workex          1  110.159 138.16
## - degree_p        1  112.815 140.81
## - mba_p           1  115.501 143.50
## - ssc_p           1  140.508 168.51
##
## Step:  AIC=127.33
## status ~ gender + ssc_p + ssc_b + hsc_p + hsc_b + degree_p +
##     degree_t + etest_p + mba_p + specialisation + workex
##
##                  Df Deviance    AIC
## - ssc_b           1   101.42 125.42
## - specialisation  1   101.61 125.61
## - hsc_b           1   101.62 125.62
## - etest_p         1   101.78 125.78
```

```
## <none>              101.33 127.33
## - gender          1   104.70 128.70
## - degree_t        2   106.94 128.94
## - hsc_p           1   110.96 134.96
## - workex          1   114.02 138.02
## - degree_p        1   114.89 138.89
## - mba_p           1   117.06 141.06
## - ssc_p           1   146.89 170.89
##
## Step:  AIC=125.42
## status ~ gender + ssc_p + hsc_p + hsc_b + degree_p + degree_t +
##     etest_p + mba_p + specialisation + workex
##
##                  Df Deviance    AIC
## - specialisation  1   101.78 123.78
## - etest_p         1   102.06 124.06
## - hsc_b           1   102.19 124.19
## <none>              101.42 125.42
## - gender          1   105.09 127.09
## - degree_t        2   107.29 127.29
## - hsc_p           1   111.12 133.12
## - workex          1   114.07 136.07
## - degree_p        1   115.14 137.15
## - mba_p           1   117.11 139.11
## - ssc_p           1   151.14 173.14
##
## Step:  AIC=123.78
## status ~ gender + ssc_p + hsc_p + hsc_b + degree_p + degree_t +
##     etest_p + mba_p + workex
##
##            Df Deviance    AIC
## - etest_p   1   102.35 122.35
## - hsc_b     1   102.50 122.50
## <none>         101.78 123.78
## - degree_t  2   108.08 126.08
## - gender    1   106.28 126.28
## - hsc_p     1   112.21 132.21
## - workex    1   115.68 135.68
## - degree_p  1   116.78 136.78
## - mba_p     1   117.58 137.58
## - ssc_p     1   151.43 171.43
##
## Step:  AIC=122.35
## status ~ gender + ssc_p + hsc_p + hsc_b + degree_p + degree_t +
##     mba_p + workex
##
##            Df Deviance    AIC
## - hsc_b     1   102.98 120.98
## <none>         102.35 122.35
## - degree_t  2   108.51 124.51
## - gender    1   106.61 124.61
## - hsc_p     1   113.53 131.53
## - workex    1   116.43 134.43
## - degree_p  1   117.08 135.08
```

```
## - mba_p     1   118.76 136.76
## - ssc_p     1   151.47 169.47
##
## Step:  AIC=120.98
## status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex
##
##            Df Deviance    AIC
## <none>          102.98 120.98
## - degree_t  2   108.78 122.78
## - gender    1   108.15 124.15
## - hsc_p     1   113.76 129.76
## - workex    1   116.94 132.94
## - degree_p  1   117.57 133.57
## - mba_p     1   118.77 134.77
## - ssc_p     1   151.75 167.75
```

```
summary(fit2)
```

```
##
## Call:
## glm(formula = status ~ gender + ssc_p + hsc_p + degree_p + degree_t +
##     mba_p + workex, family = binomial, data = Placement_Data_Full_Class)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.52339  -0.13461   0.06683   0.33202   2.34615
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -20.89182    5.01424  -4.166 3.09e-05 ***
## genderM             1.37280    0.62876   2.183 0.029009 *
## ssc_p               0.22242    0.04177   5.325 1.01e-07 ***
## hsc_p               0.09953    0.03406   2.922 0.003473 **
## degree_p            0.18706    0.05375   3.480 0.000501 ***
## degree_tOthers     -0.47685    1.16112  -0.411 0.681304
## degree_tSci&Tech   -1.42573    0.61123  -2.333 0.019672 *
## mba_p              -0.19628    0.05227  -3.755 0.000173 ***
## workexYes           2.29036    0.69033   3.318 0.000907 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 266.77  on 214  degrees of freedom
## Residual deviance: 102.98  on 206  degrees of freedom
## AIC: 120.98
##
## Number of Fisher Scoring iterations: 7
```

After doing the stepwise variable selection method using AIC, the signi???cant variables are:

1. gender

2. ssc_p
3. hsc_p
4. degree_p
5. degree_t
6. mba_p
7. workex

From the value of this coefficients, we can see that mba mark is negative as -0.19628 whereas the degree mark is positive as 0.18706 we start to grow this hypothesis that help us explain this observation in this data set that maybe this placement is taken in the time during their mba period, so the more time they use to study, the less time and effort and energy they use for placement, so there is likely they could not find one, whereas for students who results not that good in their mba, they must use more time to find placement so their likely to get one placement during this mba period. and this less effort in study because they cannot carry results and placement together.

now, let's trying to work with these 7 variables from now on to look a bit closer for all the factors

```
fit1<- glm(status~ gender+ssc_p +hsc_p + degree_p + degree_t+
                   mba_p+workex, Placement_Data_Full_Class,family=binomial)

fit2<- glm(status~ gender+ssc_p +hsc_p + degree_p + degree_t+
                   mba_p+workex+ ssc_p*hsc_p*degree_p*mba_p,
           Placement_Data_Full_Class,family=binomial)
summary(fit1)
```

```
##
## Call:
## glm(formula = status ~ gender + ssc_p + hsc_p + degree_p + degree_t +
##     mba_p + workex, family = binomial, data = Placement_Data_Full_Class)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.52339  -0.13461   0.06683   0.33202   2.34615
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -20.89182    5.01424  -4.166 3.09e-05 ***
## genderM            1.37280    0.62876   2.183 0.029009 *
## ssc_p              0.22242    0.04177   5.325 1.01e-07 ***
## hsc_p              0.09953    0.03406   2.922 0.003473 **
## degree_p           0.18706    0.05375   3.480 0.000501 ***
## degree_tOthers    -0.47685    1.16112  -0.411 0.681304
## degree_tSci&Tech  -1.42573    0.61123  -2.333 0.019672 *
## mba_p             -0.19628    0.05227  -3.755 0.000173 ***
## workexYes          2.29036    0.69033   3.318 0.000907 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 266.77  on 214  degrees of freedom
## Residual deviance: 102.98  on 206  degrees of freedom
## AIC: 120.98
##
```

```
## Number of Fisher Scoring iterations: 7
```

```r
summary(fit2)
```

```
##
## Call:
## glm(formula = status ~ gender + ssc_p + hsc_p + degree_p + degree_t +
##     mba_p + workex + ssc_p * hsc_p * degree_p * mba_p, family = binomial,
##     data = Placement_Data_Full_Class)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.72018  -0.18974   0.04718   0.22180   2.45967
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 1.351e+03  1.988e+03   0.679 0.496921
## genderM                     1.198e+00  6.807e-01   1.760 0.078442 .
## ssc_p                      -1.624e+01  3.062e+01  -0.530 0.595873
## hsc_p                      -2.852e+01  3.234e+01  -0.882 0.377817
## degree_p                   -2.337e+01  3.048e+01  -0.767 0.443332
## degree_tOthers             -2.897e-01  1.350e+00  -0.215 0.830082
## degree_tSci&Tech           -1.188e+00  7.160e-01  -1.659 0.097146 .
## mba_p                      -2.067e+01  3.183e+01  -0.649 0.516181
## workexYes                   2.801e+00  7.772e-01   3.603 0.000314 ***
## ssc_p:hsc_p                 3.751e-01  4.995e-01   0.751 0.452636
## ssc_p:degree_p              2.816e-01  4.654e-01   0.605 0.545115
## hsc_p:degree_p              4.634e-01  4.947e-01   0.937 0.348871
## ssc_p:mba_p                 2.377e-01  4.829e-01   0.492 0.622499
## hsc_p:mba_p                 4.284e-01  5.093e-01   0.841 0.400325
## degree_p:mba_p              3.631e-01  4.880e-01   0.744 0.456858
## ssc_p:hsc_p:degree_p       -5.967e-03  7.575e-03  -0.788 0.430816
## ssc_p:hsc_p:mba_p          -5.505e-03  7.735e-03  -0.712 0.476626
## ssc_p:degree_p:mba_p       -4.250e-03  7.330e-03  -0.580 0.562042
## hsc_p:degree_p:mba_p       -7.095e-03  7.786e-03  -0.911 0.362159
## ssc_p:hsc_p:degree_p:mba_p  8.997e-05  1.171e-04   0.769 0.442189
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 266.771  on 214  degrees of freedom
## Residual deviance:  92.511  on 195  degrees of freedom
## AIC: 132.51
##
## Number of Fisher Scoring iterations: 8
```

```r
anova(fit1,fit2,test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex
```

```
## Model 2: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex + ssc_p * hsc_p * degree_p * mba_p
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       206    102.982
## 2       195     92.511 11   10.471   0.4886
```

```
anova(fit1,fit2,test= "Rao")
```

```
## Analysis of Deviance Table
##
## Model 1: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex
## Model 2: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex + ssc_p * hsc_p * degree_p * mba_p
##   Resid. Df Resid. Dev Df Deviance    Rao Pr(>Chi)
## 1       206    102.982
## 2       195     92.511 11   10.471 9.4789   0.5778
```

```
wald(fit2,":")
```

```
##   numDF denDF   F-value p-value
## :    11   195 0.7099231 0.72827
##                           Estimate  Std.Error DF  t-value   p-value Lower 0.95
## ssc_p:hsc_p                0.375123 0.499480  195  0.751028 0.45354 -0.609953
## ssc_p:degree_p             0.281584 0.465354  195  0.605096 0.54582 -0.636189
## hsc_p:degree_p             0.463431 0.494706  195  0.936782 0.35003 -0.512229
## ssc_p:mba_p                0.237722 0.482868  195  0.492311 0.62305 -0.714593
## hsc_p:mba_p                0.428380 0.509345  195  0.841042 0.40135 -0.576152
## degree_p:mba_p             0.363097 0.488013  195  0.744031 0.45775 -0.599364
## ssc_p:hsc_p:degree_p      -0.005967 0.007575  195 -0.787795 0.43177 -0.020906
## ssc_p:hsc_p:mba_p         -0.005505 0.007735  195 -0.711740 0.47748 -0.020760
## ssc_p:degree_p:mba_p      -0.004250 0.007330  195 -0.579812 0.56271 -0.018705
## hsc_p:degree_p:mba_p      -0.007095 0.007786  195 -0.911259 0.36328 -0.022450
## ssc_p:hsc_p:degree_p:mba_p  0.000090 0.000117  195  0.768502 0.44312 -0.000141
##                           Upper 0.95
## ssc_p:hsc_p               1.360199
## ssc_p:degree_p            1.199357
## hsc_p:degree_p            1.439092
## ssc_p:mba_p               1.190037
## hsc_p:mba_p               1.432912
## degree_p:mba_p            1.325558
## ssc_p:hsc_p:degree_p      0.008972
## ssc_p:hsc_p:mba_p         0.009750
## ssc_p:degree_p:mba_p      0.010206
## hsc_p:degree_p:mba_p      0.008260
## ssc_p:hsc_p:degree_p:mba_p 0.000321
```

Puting all the percentage results of their high school to univeisty to mba together, in fit2 model, only work experience is significant the p value for LRT = 0.4886 is very big shows this model is not that significant overall

```r
fit3<- glm(status~ gender+ssc_p +hsc_p + degree_p + degree_t+
                   mba_p+workex, Placement_Data_Full_Class,family=binomial)
fit4<- glm(status~ gender*(ssc_p +hsc_p + degree_p + degree_t+
                   mba_p+workex),
           Placement_Data_Full_Class,family=binomial)

summary(fit3)
```

```
##
## Call:
## glm(formula = status ~ gender + ssc_p + hsc_p + degree_p + degree_t +
##     mba_p + workex, family = binomial, data = Placement_Data_Full_Class)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.52339  -0.13461   0.06683   0.33202   2.34615
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -20.89182    5.01424  -4.166 3.09e-05 ***
## genderM             1.37280    0.62876   2.183 0.029009 *
## ssc_p               0.22242    0.04177   5.325 1.01e-07 ***
## hsc_p               0.09953    0.03406   2.922 0.003473 **
## degree_p            0.18706    0.05375   3.480 0.000501 ***
## degree_tOthers     -0.47685    1.16112  -0.411 0.681304
## degree_tSci&Tech   -1.42573    0.61123  -2.333 0.019672 *
## mba_p              -0.19628    0.05227  -3.755 0.000173 ***
## workexYes           2.29036    0.69033   3.318 0.000907 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 266.77  on 214  degrees of freedom
## Residual deviance: 102.98  on 206  degrees of freedom
## AIC: 120.98
##
## Number of Fisher Scoring iterations: 7
```

```r
summary(fit4)
```

```
##
## Call:
## glm(formula = status ~ gender * (ssc_p + hsc_p + degree_p + degree_t +
##     mba_p + workex), family = binomial, data = Placement_Data_Full_Class)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.86267  -0.13697   0.05543   0.24744   2.09990
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)               -12.35629    7.72311  -1.600 0.109618
## genderM                   -14.15873   10.27801  -1.378 0.168335
## ssc_p                       0.25451    0.07541   3.375 0.000738 ***
## hsc_p                       0.18712    0.08244   2.270 0.023226 *
## degree_p                    0.16949    0.08550   1.982 0.047455 *
## degree_tOthers             -0.10559    2.78742  -0.038 0.969784
## degree_tSci&Tech           -1.13007    1.26570  -0.893 0.371942
## mba_p                      -0.43389    0.14122  -3.072 0.002124 **
## workexYes                   3.32147    1.51170   2.197 0.028008 *
## genderM:ssc_p              -0.04004    0.09442  -0.424 0.671546
## genderM:hsc_p              -0.11688    0.09264  -1.262 0.207058
## genderM:degree_p            0.08980    0.11817   0.760 0.447338
## genderM:degree_tOthers     -2.10196    3.73749  -0.562 0.573845
## genderM:degree_tSci&Tech   -0.49918    1.47642  -0.338 0.735286
## genderM:mba_p               0.32123    0.15615   2.057 0.039671 *
## genderM:workexYes          -1.43521    1.71618  -0.836 0.402998
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 266.771  on 214  degrees of freedom
## Residual deviance:  92.767  on 199  degrees of freedom
## AIC: 124.77
##
## Number of Fisher Scoring iterations: 7
```

```
anova(fit3,fit4,test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex
## Model 2: status ~ gender * (ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       206    102.982
## 2       199     92.767  7   10.215   0.1767
```

```
anova(fit3,fit4,test= "Rao")
```

```
## Analysis of Deviance Table
##
## Model 1: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex
## Model 2: status ~ gender * (ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex)
##   Resid. Df Resid. Dev Df Deviance    Rao Pr(>Chi)
## 1       206    102.982
## 2       199     92.767  7   10.215 9.2473   0.2354
```

```r
wald(fit4,":")
```

```
##   numDF denDF  F-value p-value
## :     7   199 1.095967 0.36715
##                          Estimate  Std.Error DF  t-value   p-value Lower 0.95
## genderM:ssc_p            -0.040036 0.094418  199 -0.424028 0.67200 -0.226224
## genderM:hsc_p            -0.116879 0.092636  199 -1.261697 0.20853 -0.299553
## genderM:degree_p          0.089796 0.118174  199  0.759860 0.44824 -0.143239
## genderM:degree_tOthers   -2.101956 3.737485  199 -0.562399 0.57448 -9.472115
## genderM:degree_tSci&Tech -0.499180 1.476418  199 -0.338102 0.73564 -3.410613
## genderM:mba_p             0.321227 0.156151  199  2.057156 0.04098  0.013304
## genderM:workexYes        -1.435207 1.716182  199 -0.836279 0.40400 -4.819444
##                          Upper 0.95
## genderM:ssc_p            0.146152
## genderM:hsc_p            0.065796
## genderM:degree_p         0.322831
## genderM:degree_tOthers   5.268202
## genderM:degree_tSci&Tech 2.412253
## genderM:mba_p            0.629151
## genderM:workexYes        1.949030
```

Putting gender with interaction with all the school results and prior work experience together, in fit4 model, only work experience is significant the p value for LRT = 0.1767 is much small though still not significant to show this model is significant overall, but we can see gender place a imporant role here.

```r
fit5<- glm(status~ gender+ssc_p +hsc_p + degree_p + degree_t+
                   mba_p+workex, Placement_Data_Full_Class,family=binomial)
fit6<- glm(status~ gender+ssc_p +hsc_p + degree_p + degree_t+
                   mba_p+workex+ degree_p*degree_t,
           Placement_Data_Full_Class,family=binomial)


summary(fit5)
```

```
##
## Call:
## glm(formula = status ~ gender + ssc_p + hsc_p + degree_p + degree_t +
##     mba_p + workex, family = binomial, data = Placement_Data_Full_Class)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.52339  -0.13461   0.06683   0.33202   2.34615
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -20.89182    5.01424  -4.166 3.09e-05 ***
## genderM             1.37280    0.62876   2.183 0.029009 *
## ssc_p               0.22242    0.04177   5.325 1.01e-07 ***
## hsc_p               0.09953    0.03406   2.922 0.003473 **
## degree_p            0.18706    0.05375   3.480 0.000501 ***
## degree_tOthers     -0.47685    1.16112  -0.411 0.681304
## degree_tSci&Tech   -1.42573    0.61123  -2.333 0.019672 *
```

```
## mba_p              -0.19628    0.05227  -3.755 0.000173 ***
## workexYes           2.29036    0.69033   3.318 0.000907 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 266.77  on 214  degrees of freedom
## Residual deviance: 102.98  on 206  degrees of freedom
## AIC: 120.98
##
## Number of Fisher Scoring iterations: 7
```

```r
summary(fit6)
```

```
##
## Call:
## glm(formula = status ~ gender + ssc_p + hsc_p + degree_p + degree_t +
##     mba_p + workex + degree_p * degree_t, family = binomial,
##     data = Placement_Data_Full_Class)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.49431  -0.12643   0.05114   0.33991   2.34125
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -20.83853    5.24918  -3.970 7.19e-05 ***
## genderM                     1.45636    0.64084   2.273  0.02305 *
## ssc_p                       0.22520    0.04245   5.305 1.13e-07 ***
## hsc_p                       0.10376    0.03511   2.956  0.00312 **
## degree_p                    0.17466    0.06080   2.873  0.00407 **
## degree_tOthers              7.21066   12.89399   0.559  0.57601
## degree_tSci&Tech           -7.12769    8.33132  -0.856  0.39226
## mba_p                      -0.19221    0.05263  -3.652  0.00026 ***
## workexYes                   2.26007    0.69014   3.275  0.00106 **
## degree_p:degree_tOthers    -0.13134    0.21603  -0.608  0.54321
## degree_p:degree_tSci&Tech   0.08601    0.12578   0.684  0.49410
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 266.77  on 214  degrees of freedom
## Residual deviance: 102.13  on 204  degrees of freedom
## AIC: 124.13
##
## Number of Fisher Scoring iterations: 7
```

```r
anova(fit5,fit6,test="LRT")
```

```
## Analysis of Deviance Table
##
```

```
## Model 1: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex
## Model 2: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex + degree_p * degree_t
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       206     102.98
## 2       204     102.13  2  0.85318   0.6527
```

```
anova(fit5,fit6,test= "Rao")
```

```
## Analysis of Deviance Table
##
## Model 1: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex
## Model 2: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex + degree_p * degree_t
##   Resid. Df Resid. Dev Df Deviance     Rao Pr(>Chi)
## 1       206     102.98
## 2       204     102.13  2  0.85318 0.86062   0.6503
```

```
wald(fit6,":")
```

```
##   numDF denDF   F-value p-value
## :     2   204 0.4631865 0.62994
##                           Estimate  Std.Error DF  t-value   p-value Lower 0.95
## degree_p:degree_tOthers   -0.131340 0.216031  204 -0.607970 0.54388 -0.557279
## degree_p:degree_tSci&Tech  0.086011 0.125782  204  0.683806 0.49487 -0.161989
##                           Upper 0.95
## degree_p:degree_tOthers   0.294599
## degree_p:degree_tSci&Tech 0.334011
```

Put interaction term as the percentage results of university to university major, we can see school results are significant, work experience are significant, and gender are slightly significant in model 6 based on p value the p value for LRT = 0.6527 is very big shows this model overall is not that significant

```
fit7<- glm(status~ gender+ssc_p +hsc_p + degree_p + degree_t+
                   mba_p+workex, Placement_Data_Full_Class,family=binomial)
fit8<- glm(status~ gender+ssc_p +hsc_p + degree_p + degree_t+
                   mba_p+workex+ gender*degree_t,
           Placement_Data_Full_Class,family=binomial)


summary(fit7)
```

```
##
## Call:
## glm(formula = status ~ gender + ssc_p + hsc_p + degree_p + degree_t +
##     mba_p + workex, family = binomial, data = Placement_Data_Full_Class)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.52339  -0.13461   0.06683   0.33202   2.34615
```

```
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -20.89182    5.01424  -4.166 3.09e-05 ***
## genderM            1.37280    0.62876   2.183 0.029009 *
## ssc_p              0.22242    0.04177   5.325 1.01e-07 ***
## hsc_p              0.09953    0.03406   2.922 0.003473 **
## degree_p           0.18706    0.05375   3.480 0.000501 ***
## degree_tOthers    -0.47685    1.16112  -0.411 0.681304
## degree_tSci&Tech  -1.42573    0.61123  -2.333 0.019672 *
## mba_p             -0.19628    0.05227  -3.755 0.000173 ***
## workexYes          2.29036    0.69033   3.318 0.000907 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 266.77  on 214  degrees of freedom
## Residual deviance: 102.98  on 206  degrees of freedom
## AIC: 120.98
##
## Number of Fisher Scoring iterations: 7
```

```
summary(fit8)
```

```
##
## Call:
## glm(formula = status ~ gender + ssc_p + hsc_p + degree_p + degree_t +
##     mba_p + workex + gender * degree_t, family = binomial, data = Placement_Data_Full_Class)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5380  -0.1332   0.0614   0.3352   2.3550
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -20.92055    4.99618  -4.187 2.82e-05 ***
## genderM                    1.38615    0.72907   1.901 0.057267 .
## ssc_p                      0.22236    0.04245   5.239 1.62e-07 ***
## hsc_p                      0.09348    0.03520   2.655 0.007920 **
## degree_p                   0.19180    0.05379   3.566 0.000363 ***
## degree_tOthers             0.48073    1.53648   0.313 0.754375
## degree_tSci&Tech          -1.78967    1.02947  -1.738 0.082133 .
## mba_p                     -0.19442    0.05200  -3.739 0.000185 ***
## workexYes                  2.29769    0.70337   3.267 0.001088 **
## genderM:degree_tOthers    -2.91353    2.88893  -1.009 0.313207
## genderM:degree_tSci&Tech   0.47973    1.19816   0.400 0.688873
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 266.77  on 214  degrees of freedom
## Residual deviance: 101.47  on 204  degrees of freedom
```

```
## AIC: 123.47
##
## Number of Fisher Scoring iterations: 7
```

```
anova(fit7,fit8,test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex
## Model 2: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex + gender * degree_t
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       206     102.98
## 2       204     101.47  2   1.5112   0.4697
```

```
anova(fit7,fit8,test= "Rao")
```

```
## Analysis of Deviance Table
##
## Model 1: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex
## Model 2: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex + gender * degree_t
##   Resid. Df Resid. Dev Df Deviance    Rao Pr(>Chi)
## 1       206     102.98
## 2       204     101.47  2   1.5112 1.3237   0.5159
```

```
wald(fit8,":")
```

```
##   numDF denDF   F-value p-value
## :     2   204 0.6438146 0.52635
##                          Estimate  Std.Error DF  t-value   p-value Lower 0.95
## genderM:degree_tOthers   -2.913530 2.888926  204 -1.008517 0.31440 -8.609512
## genderM:degree_tSci&Tech  0.479726 1.198162  204  0.400385 0.68929 -1.882644
##                          Upper 0.95
## genderM:degree_tOthers   2.782452
## genderM:degree_tSci&Tech 2.842096
```

Put interaction for the gender and their university major,we can see school results are significant, work experience are significant in model 8 based on p value, but the choice of degree does not seem that important here. the p value for LRT = 0.4697 is very big shows this model overall is not that significant

```
fit9<- glm(status~ gender+ssc_p +hsc_p + degree_p + degree_t+
                   mba_p+workex, Placement_Data_Full_Class,family=binomial)
fit10<- glm(status~ gender+ssc_p +hsc_p + degree_p + degree_t+
                   mba_p+workex+ ssc_p*hsc_p*degree_p*mba_p*workex,
           Placement_Data_Full_Class,family=binomial)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
summary(fit9)
```

```
##
## Call:
## glm(formula = status ~ gender + ssc_p + hsc_p + degree_p + degree_t +
##     mba_p + workex, family = binomial, data = Placement_Data_Full_Class)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.52339  -0.13461   0.06683   0.33202   2.34615
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -20.89182    5.01424  -4.166 3.09e-05 ***
## genderM              1.37280    0.62876   2.183 0.029009 *
## ssc_p                0.22242    0.04177   5.325 1.01e-07 ***
## hsc_p                0.09953    0.03406   2.922 0.003473 **
## degree_p             0.18706    0.05375   3.480 0.000501 ***
## degree_tOthers      -0.47685    1.16112  -0.411 0.681304
## degree_tSci&Tech    -1.42573    0.61123  -2.333 0.019672 *
## mba_p               -0.19628    0.05227  -3.755 0.000173 ***
## workexYes            2.29036    0.69033   3.318 0.000907 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 266.77  on 214  degrees of freedom
## Residual deviance: 102.98  on 206  degrees of freedom
## AIC: 120.98
##
## Number of Fisher Scoring iterations: 7
```

```r
summary(fit10)
```

```
##
## Call:
## glm(formula = status ~ gender + ssc_p + hsc_p + degree_p + degree_t +
##     mba_p + workex + ssc_p * hsc_p * degree_p * mba_p * workex,
##     family = binomial, data = Placement_Data_Full_Class)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.74330  -0.05187   0.00000   0.16144   2.61436
##
## Coefficients:
##                                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)                            2.584e+03  2.122e+03   1.218   0.2232
## genderM                                1.443e+00  8.388e-01   1.720   0.0854 .
## ssc_p                                 -3.261e+01  2.986e+01  -1.092   0.2747
## hsc_p                                 -5.441e+01  3.660e+01  -1.487   0.1371
## degree_p                              -4.038e+01  3.342e+01  -1.208   0.2270
```

```
## degree_tOthers                      -2.480e+00  3.809e+00  -0.651   0.5149
## degree_tSci&Tech                    -1.687e+00  8.753e-01  -1.927   0.0540 .
## mba_p                               -4.159e+01  3.729e+01  -1.115   0.2648
## workexYes                            8.350e+04  1.218e+08   0.001   0.9995
## ssc_p:hsc_p                          7.207e-01  5.162e-01   1.396   0.1627
## ssc_p:degree_p                       5.040e-01  4.634e-01   1.088   0.2768
## hsc_p:degree_p                       8.378e-01  5.742e-01   1.459   0.1445
## ssc_p:mba_p                          5.124e-01  5.143e-01   0.996   0.3191
## hsc_p:mba_p                          8.670e-01  6.261e-01   1.385   0.1662
## degree_p:mba_p                       6.501e-01  5.892e-01   1.103   0.2699
## ssc_p:workexYes                     -9.671e+02  2.300e+06   0.000   0.9997
## hsc_p:workexYes                     -1.375e+03  1.526e+06  -0.001   0.9993
## degree_p:workexYes                  -1.677e+03  1.872e+06  -0.001   0.9993
## mba_p:workexYes                     -1.313e+03  1.867e+06  -0.001   0.9994
## ssc_p:hsc_p:degree_p                -1.094e-02  7.987e-03  -1.370   0.1707
## ssc_p:hsc_p:mba_p                   -1.132e-02  8.620e-03  -1.313   0.1891
## ssc_p:degree_p:mba_p                -7.952e-03  8.024e-03  -0.991   0.3216
## hsc_p:degree_p:mba_p                -1.341e-02  9.851e-03  -1.361   0.1736
## ssc_p:hsc_p:workexYes                1.777e+01  2.799e+04   0.001   0.9995
## ssc_p:degree_p:workexYes             2.151e+01  3.536e+04   0.001   0.9995
## hsc_p:degree_p:workexYes             2.635e+01  2.335e+04   0.001   0.9991
## ssc_p:mba_p:workexYes                1.415e+01  3.431e+04   0.000   0.9997
## hsc_p:mba_p:workexYes                2.182e+01  2.412e+04   0.001   0.9993
## degree_p:mba_p:workexYes             2.661e+01  2.854e+04   0.001   0.9993
## ssc_p:hsc_p:degree_p:mba_p           1.731e-04  1.339e-04   1.292   0.1963
## ssc_p:hsc_p:degree_p:workexYes      -3.601e-01  4.297e+02  -0.001   0.9993
## ssc_p:hsc_p:mba_p:workexYes         -2.685e-01  4.264e+02  -0.001   0.9995
## ssc_p:degree_p:mba_p:workexYes      -3.245e-01  5.252e+02  -0.001   0.9995
## hsc_p:degree_p:mba_p:workexYes      -4.220e-01  3.666e+02  -0.001   0.9991
## ssc_p:hsc_p:degree_p:mba_p:workexYes  5.549e-03  6.508e+00   0.001   0.9993
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 266.771  on 214  degrees of freedom
## Residual deviance:  69.908  on 180  degrees of freedom
## AIC: 139.91
##
## Number of Fisher Scoring iterations: 22
```

```
anova(fit9,fit10,test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex
## Model 2: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex + ssc_p * hsc_p * degree_p * mba_p * workex
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       206    102.982
## 2       180     69.908 26   33.074     0.16
```

```r
anova(fit9,fit10,test= "Rao")
```

```
## Analysis of Deviance Table
##
## Model 1: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex
## Model 2: status ~ gender + ssc_p + hsc_p + degree_p + degree_t + mba_p +
##     workex + ssc_p * hsc_p * degree_p * mba_p * workex
##   Resid. Df Resid. Dev Df Deviance  Rao Pr(>Chi)
## 1       206    102.982
## 2       180     69.908 26   33.074 24.2   0.5645
```

```r
#wald(fit10,":")
```

Put the work experience with results they have, there is no significant for any individual in model 10, the p value for LRT = 0.16 is getting smaller, although this model is not that significant overall, but it might means that work experince is somewhat important to have

Deep analysis

after this try out of type I tests we starts to learn that work experience, gender, and degree_p (results in their univeristy) have a big say in the overal model, so i am going to dig a little bit to test out the data with these varibales.

```r
fita<- glm(status~gender+ degree_p +workex, Placement_Data_Full_Class,family=binomial)


fitb<- glm(status~gender+ degree_p*workex, Placement_Data_Full_Class,family=binomial)

summary(fita)
```

```
##
## Call:
## glm(formula = status ~ gender + degree_p + workex, family = binomial,
##     data = Placement_Data_Full_Class)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3909  -0.7244   0.3767   0.6844   2.0711
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -13.75421    2.27933  -6.034 1.60e-09 ***
## genderM       0.93389    0.38604   2.419   0.0156 *
## degree_p      0.20954    0.03426   6.116 9.62e-10 ***
## workexYes     1.39148    0.43834   3.174   0.0015 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 266.77  on 214  degrees of freedom
## Residual deviance: 188.99  on 211  degrees of freedom
```

```
## AIC: 196.99
##
## Number of Fisher Scoring iterations: 5
```

```
summary(fitb)
```

```
##
## Call:
## glm(formula = status ~ gender + degree_p * workex, family = binomial,
##     data = Placement_Data_Full_Class)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4171  -0.7450   0.3505   0.7066   2.0243
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -12.99425    2.48827  -5.222 1.77e-07 ***
## genderM             0.90974    0.38856   2.341   0.0192 *
## degree_p            0.19792    0.03747   5.282 1.28e-07 ***
## workexYes          -2.27652    5.58830  -0.407   0.6837
## degree_p:workexYes  0.05824    0.08880   0.656   0.5119
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 266.77  on 214  degrees of freedom
## Residual deviance: 188.53  on 210  degrees of freedom
## AIC: 198.53
##
## Number of Fisher Scoring iterations: 6
```

```
anova(fita,fitb,test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: status ~ gender + degree_p + workex
## Model 2: status ~ gender + degree_p * workex
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       211     188.99
## 2       210     188.53  1   0.4597   0.4978
```

with the interaction of degree_p and workex, degree_p is significant, but Overall LRT = 0.4978 is not significant

```
fitc<- glm(status~gender*degree_p+ workex, Placement_Data_Full_Class,family=binomial)
summary(fitc)
```

```
##
## Call:
## glm(formula = status ~ gender * degree_p + workex, family = binomial,
```

```
##     data = Placement_Data_Full_Class)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7515  -0.6707   0.3136   0.6809   1.9239
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -8.57448    2.71383  -3.160  0.00158 **
## genderM            -10.03639    4.48311  -2.239  0.02517 *
## degree_p             0.13111    0.04068   3.223  0.00127 **
## workexYes            1.34191    0.43872   3.059  0.00222 **
## genderM:degree_p     0.17123    0.07021   2.439  0.01473 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 266.77  on 214  degrees of freedom
## Residual deviance: 182.74  on 210  degrees of freedom
## AIC: 192.74
##
## Number of Fisher Scoring iterations: 6
```

```
anova(fita,fitc,test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: status ~ gender + degree_p + workex
## Model 2: status ~ gender * degree_p + workex
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       211     188.99
## 2       210     182.74  1    6.254  0.01239 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

with the interaction of degree_p and gender, all is significant, and Overall LRT = 0.01239 is significant

```
fitd<- glm(status~gender*workex+degree_p, Placement_Data_Full_Class,family=binomial)
summary(fitd)
```

```
##
## Call:
## glm(formula = status ~ gender * workex + degree_p, family = binomial,
##     data = Placement_Data_Full_Class)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3818  -0.7097   0.3829   0.6940   2.0930
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)       -13.87478    2.31486  -5.994 2.05e-09 ***
## genderM             1.00140    0.43759   2.288   0.0221 *
## workexYes           1.59674    0.76640   2.083   0.0372 *
## degree_p            0.21077    0.03455   6.100 1.06e-09 ***
## genderM:workexYes  -0.30880    0.93332  -0.331   0.7407
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 266.77  on 214  degrees of freedom
## Residual deviance: 188.88  on 210  degrees of freedom
## AIC: 198.88
##
## Number of Fisher Scoring iterations: 5
```

```
anova(fita,fitd,test="LRT")
```

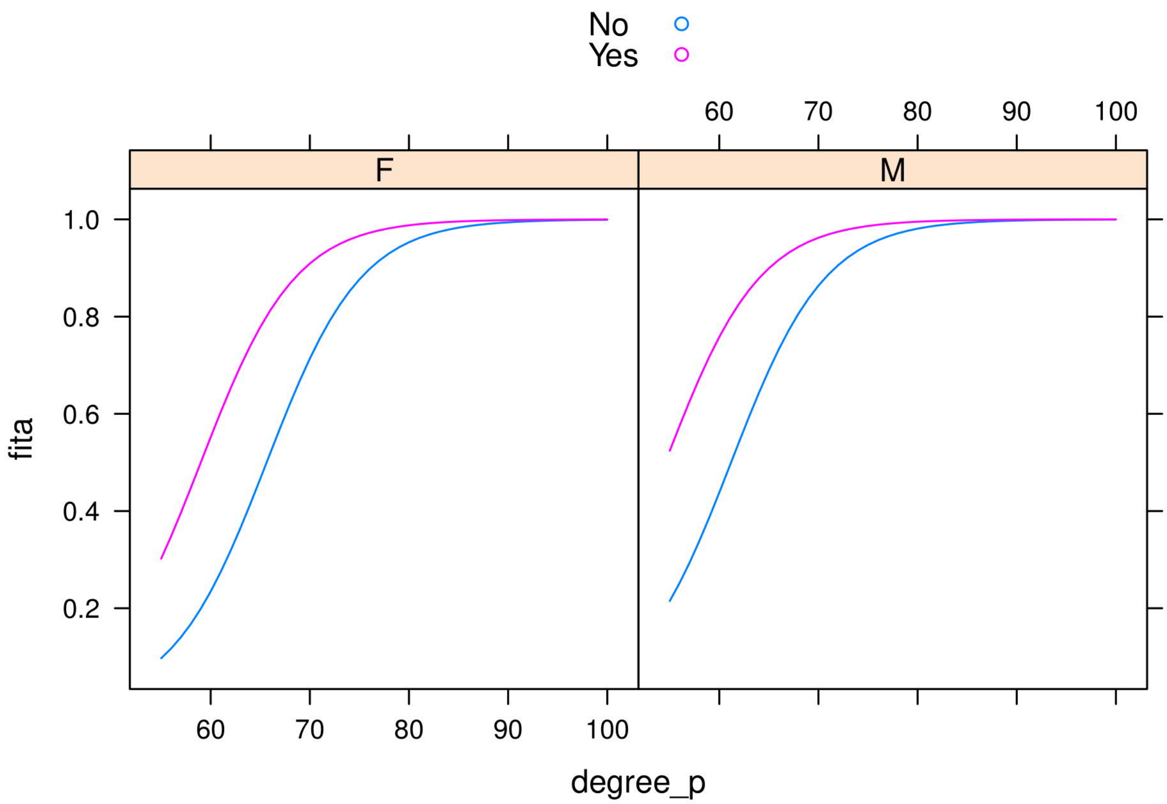
```
## Analysis of Deviance Table
##
## Model 1: status ~ gender + degree_p + workex
## Model 2: status ~ gender * workex + degree_p
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       211     188.99
## 2       210     188.88  1  0.11049   0.7396
```

with the interaction of workex and gender, all individual is significant, but interaction is not significant Overall LRT = 0.7396 is not significant

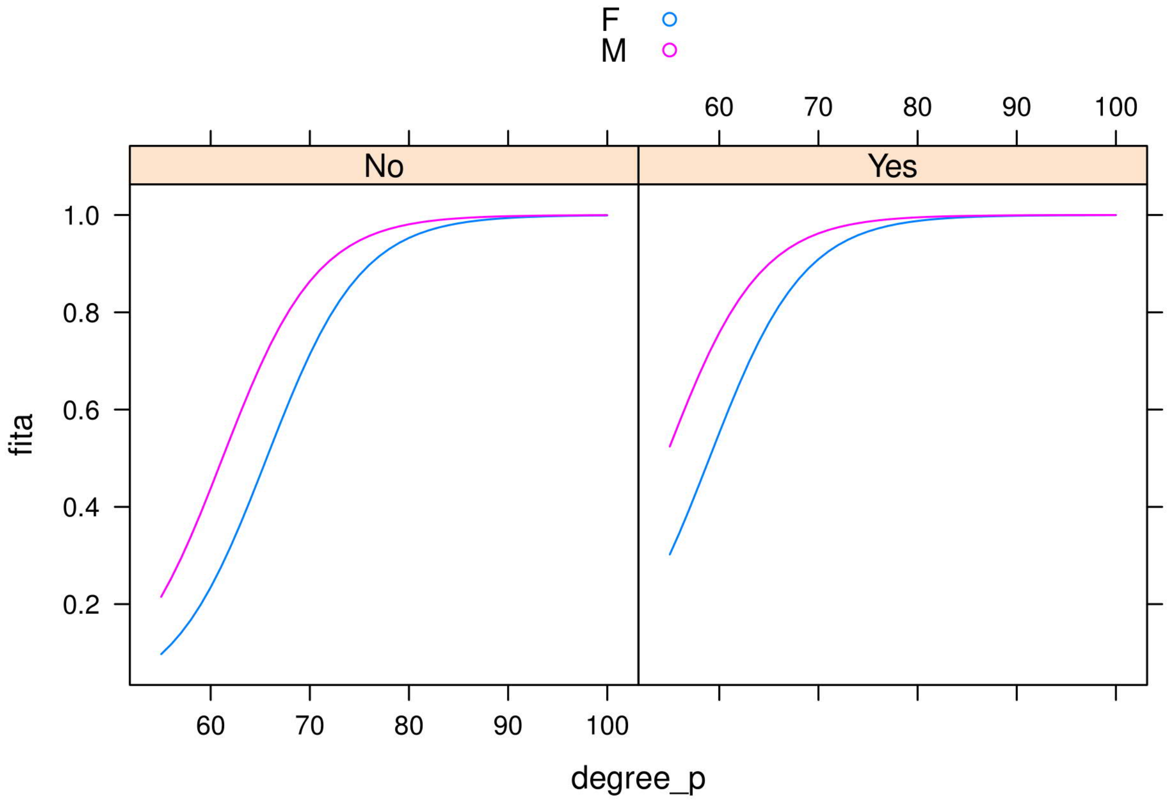
overall in this three test cases, i would say there is a strong evident that include interaction of gender and university result is a good choice.

let's try to prove it more in the predict model

```
pred<- expand.grid(gender= unique(Placement_Data_Full_Class$gender),
                   workex=unique(Placement_Data_Full_Class$workex),degree_p=seq(55,100))

pred$fita<-predict(fita,newdata=pred,type="response")
xyplot(fita~degree_p|gender,pred,groups=workex,type="l",auto.key = TRUE)
```




```r
xyplot(fita~degree_p|workex,pred,groups=gender,type="l",auto.key = TRUE)
```



the first graph shows that with a work experience, even if your grade in school not that well, you still find a placement if your have a work experience, but male have a tendency to have a lower mark with work experience to find a placement at the end. this conclusion is similar to fitc model for interaction of gender and mark.

Moreover, if you mark is well and having work experience or not is around the same as the line reach each other together.

it shows more obvious in graph 2, that for yes finding a placement, we can see male haves a higher probability to find placement with the same mark as women. this may imply we need to to find some policy to help women to find work experience so that they can be in the same fair environment to compete with man to find placement later on

#Conclusion

As we keep trying, I get this model of status~gender*degree_p+ workex to be so far the best p value for LRT =0.01239 (show significant) it shows that gender*degree_p interaction make sense on what they are. maybe women tends to do well in results as they are more diligent and can do well in exams so that their results is better. However, we also see the limitation of women need to do well than man in order to get work experience and go find a placement thus, more school resources should put on women on how to get work experience in order to narrow down the gender gap.

```
#library(car)
#library(effects)
#plot(allEffects(fita))
#avPlots(fita)
#residualPlots(fita)
```