

RESOURCE FILE PROJECT PHASE 2

Ish Durrani, Yakup Tezcan, Sophie Asselstine, Katie McCombe

GROUP 159

```
knitr::opts_chunk$set(echo = TRUE)
#Imports csv file from user file to phase2.
phase2 <- read.csv(file = "~/project.csv")
```

PROBLEM

Research Objective:

Understand the performance of individuals in prior education that determine placement in the workforce for individuals

Research Questions:

- 1.What is the mean salary offered by corporate to candidates?
- 2.Do males and females differ in their likelihood of being placed? (*Corrected*)
- 3.How does the 12th grade education percentage affect the salary of an individual?
- 4.How does the median degree percentage across students who specialized in Science, Commerce, and Arts vary?
- 5.Are students who receive an undergraduate degree in Science and Technology more likely to be placed?

SAMPLING AND STUDY DESIGN(PLAN)

The population of interest for this study are Postgraduate students. The sampling frame includes Postgraduate students who are currently attending Jain University in Bangalore. The sample for this study are 215 Marketing MBA (Postgraduate) students attending Jain University in Bangalore.

The sampling design used to obtain the sample was stratified sampling. The sampling frame was divided into 2 strata based on Post Graduation specialization: with the first being Marketing & Finance, and the second being Marketing & HR. Proportional sampling was used in order to draw a random sample of 215 students from our sampling frame. 56% of Marketing Majors from this population specialize in Finance, While 44% specialize in HR. For this reason, there was a proportional level of students from this sampling frame drawn to represent the sample, with 120/215 being Marketing and Finance, and the remaining students being Marketing and HR (95/215).

The study design used to select this sample was a survey. The variables of interest from the sample were studied observationally and information was collected from each. It is known this is an observational study because no treatment groups or controls are present in the data. The data is collected naturally and measured across a response variable. No action is made to have an effect on the outcome of the study. The information collected from each student was whether they are male or female, secondary education percentage, board

of education, higher secondary education percentage - 12th grade, and specialization in higher secondary education. In addition, degree percentage, undergraduate degree type, work experience, employability test percentage, MBA test percentage, status of placement and salary offered was collected. A survey was the appropriate study design because data was collected over several variables in order to observe the response variable, whether the student was placed or not placed.

The dataset provided did not mention or provide any sampling or study designs. As a result the sampling and study designs mentioned were made up entirely to accurately fit the data that was collected by the researchers.

SUMMARIZING AND EXPLORING DATA(DATA SECTION)

Numerical Summaries

Mean

```
#mean salary of the placed participants.  
mean(phase2$salary, na.rm=TRUE)
```

```
## [1] 288655.4
```

Variance in mean is affected by outliers or skewed data.

Median

```
#median of the employable test percentage.  
median(phase2$etest_p)
```

```
## [1] 71
```

Variance becomes apparent in median when there are outliers or skewed data.

Minimum

```
#minimum mba percentage.  
min(phase2$mba_p)
```

```
## [1] 51.21
```

Minimum values vary among the lowest values in the data.

Variance

```
#variance in the higher school education percentage.  
var(phase2$hsc_p)
```

```
## [1] 118.7557
```

Variance can vary when extreme outliers or skewed data is introduced.

Maximum

```
#maximum salary of the placed participants.  
max(phase2$salary, na.rm=TRUE)
```

```
## [1] 940000
```

Maximum values vary among the highest values in the data.

Range

```
#range of the secondary education percentage.  
range(phase2$ssc_p)
```

```
## [1] 40.89 89.40
```

Variance can be introduced by larger differences between maximum and minimum values.

IQR

```
#IQR of degree percentage.  
IQR(phase2$degree_p)
```

```
## [1] 11
```

Interquartile range varies depending on the size of the difference between the third and first quartile.

Standard deviation

```
#Standard deviation of salary of the placed participants.  
sd(phase2$salary, na.rm=TRUE)
```

```
## [1] 93457.45
```

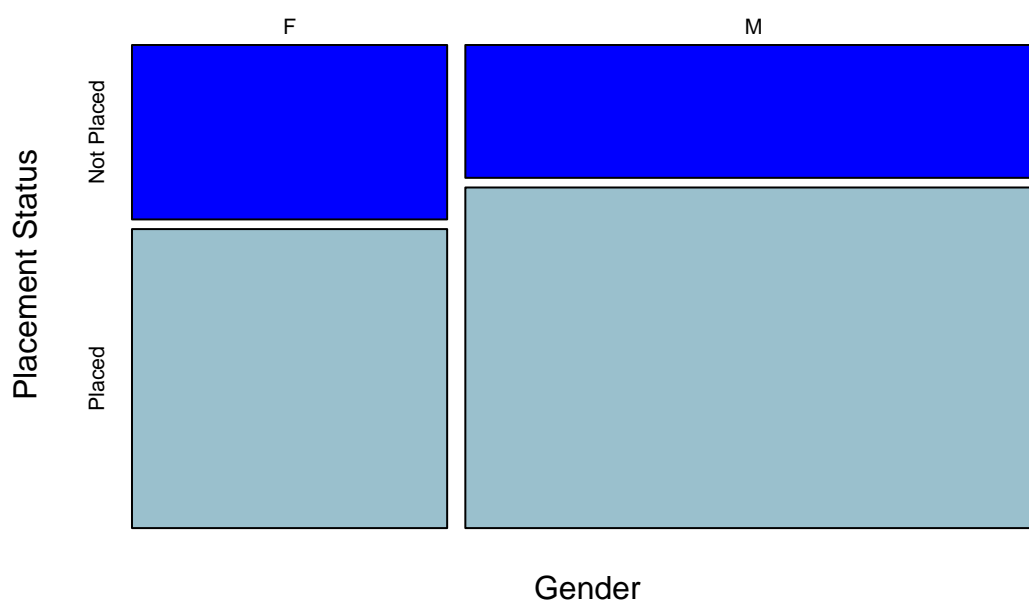
Standard deviation varies when extreme outliers or skewed data is introduced.

Graphical Summaries

Do males and females differ in their likelihood of being placed?

```
#Creates a table with gender and placement status variables.  
plot_mos <- table(phase2$gender, phase2$status)  
#Creates a mosaic plot with the table.  
mosaicplot(plot_mos, xlab="Gender", ylab="Placement Status",  
            col=c("blue", "lightblue3"),  
            main="Distribution of Gender and Placement Status")
```

Distribution of Gender and Placement Status



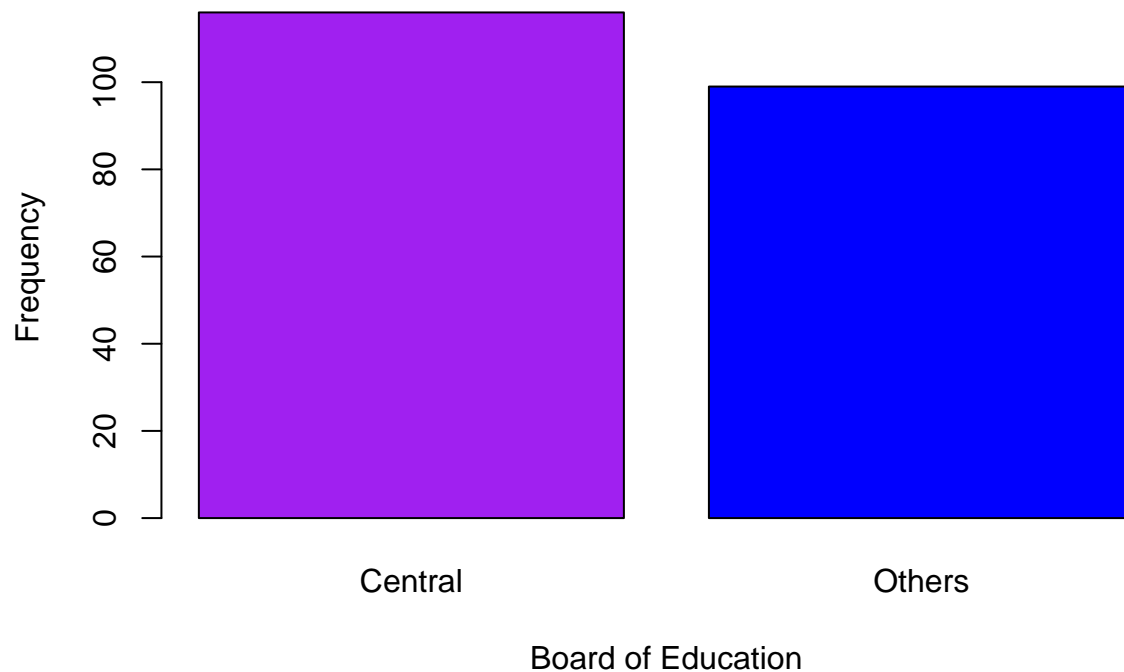
What is the proportion of participants who attended secondary school at a central board vs another board?

Bar graphs vary in whether the variables are univariate or bivariate .

Univariate

```
#Creates a table with secondary school board variable.
b_table <- table(phase2$ssc_b)
#Creates a barplot with the table.
barplot(b_table, ylab="Frequency", xlab="Board of Education", col=c("purple", "blue"),
        main="Distribution of Male and Female Students and Their School Boards")
```

Distribution of Male and Female Students and Their School Boards



Bivariate

```
#Creates a table with secondary school board and gender variables.  
b_table <- table(phase2$ssc_b, phase2$gender)  
#Creates a barplot with the table.  
barplot(b_table, ylab="Board of Education", xlab="Gender", col=c("purple", "blue"),  
        main="Distribution of Male and Female Students and Their School Boards")
```

Distribution of Male and Female Students and Their School Boards

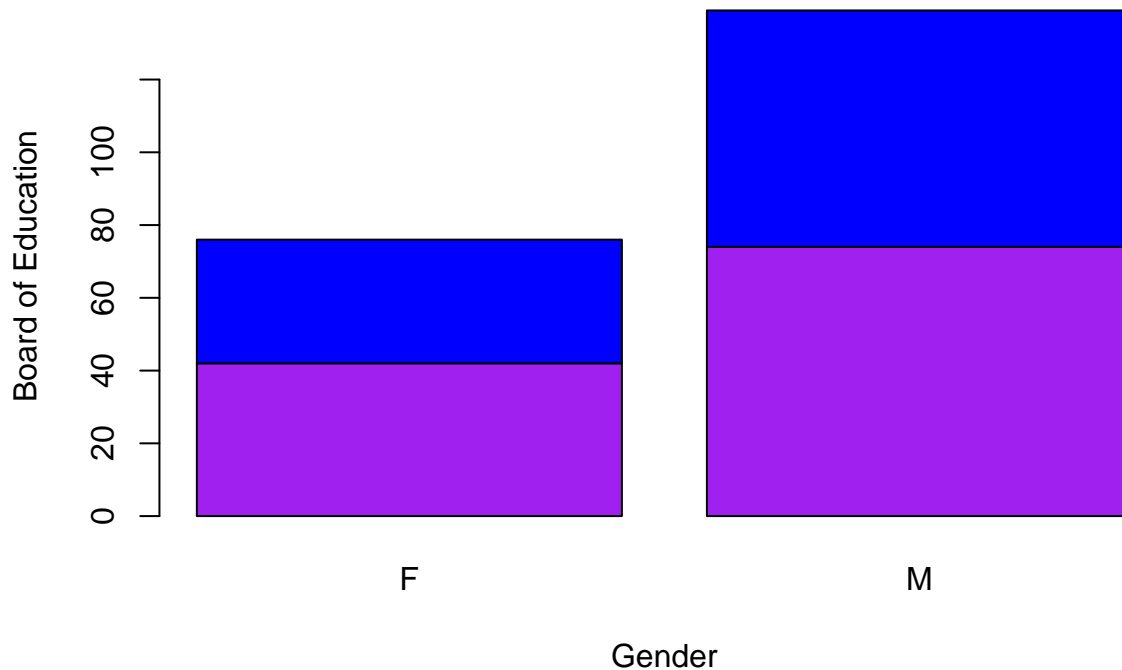
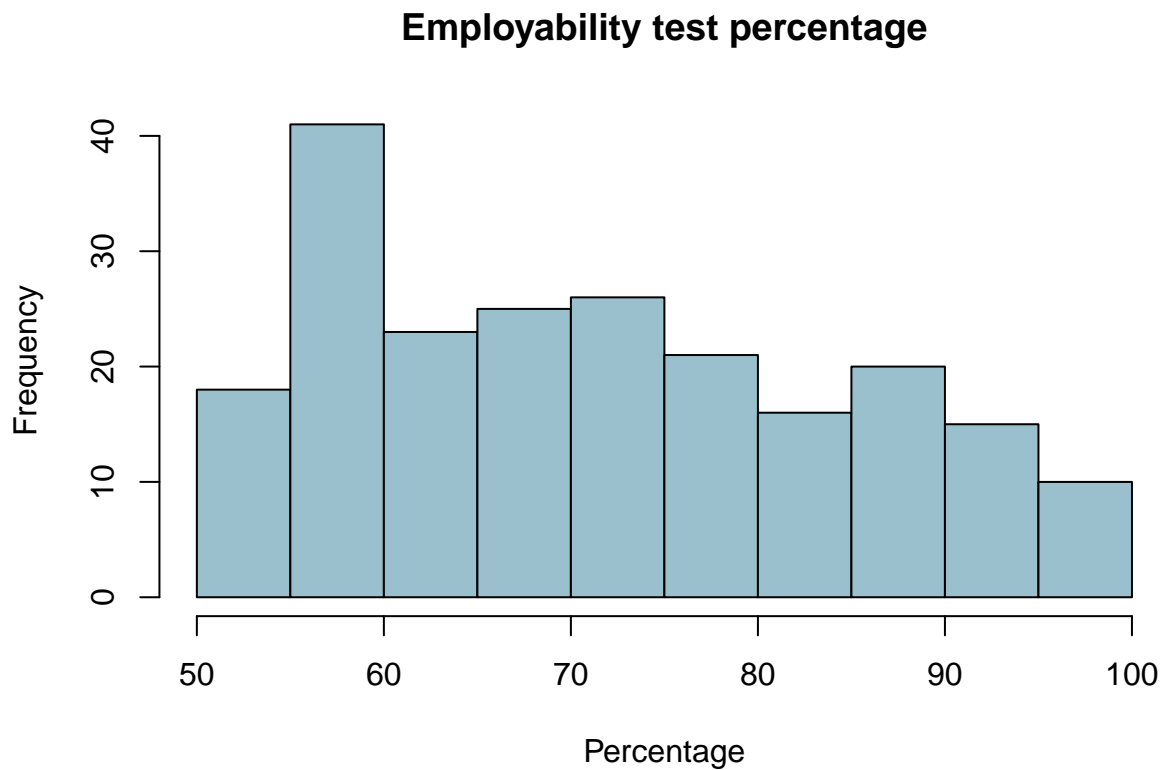


Figure 1. Frequency of participants separated by gender who attended secondary school at a central board vs another board. Purple represents students who attended a central board (n= 42 for females, n= 50 for males), blue represents students who attended another board (n= 68 for females, n= 86 for males).

What is the percentage of participants who received an Employability test score of over 80%?

```
#Creates a histogram with employability test percentage.  
hist(phase2$etest_p, xlab="Percentage", ylab="Frequency", col="lightblue3",  
      main="Employability test percentage")
```

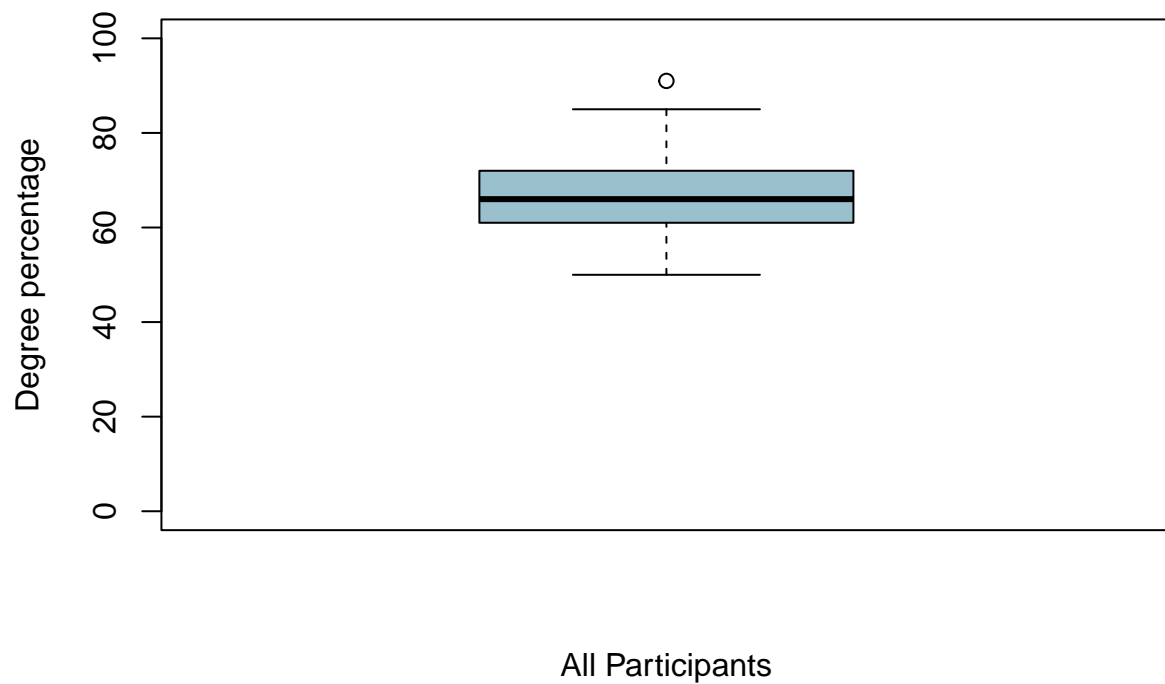


How does the median degree percentage across students who specialized in Science, Commerce, and Arts vary?

Boxplots vary in orientation, it can be vertical or horizontal.

Univariate

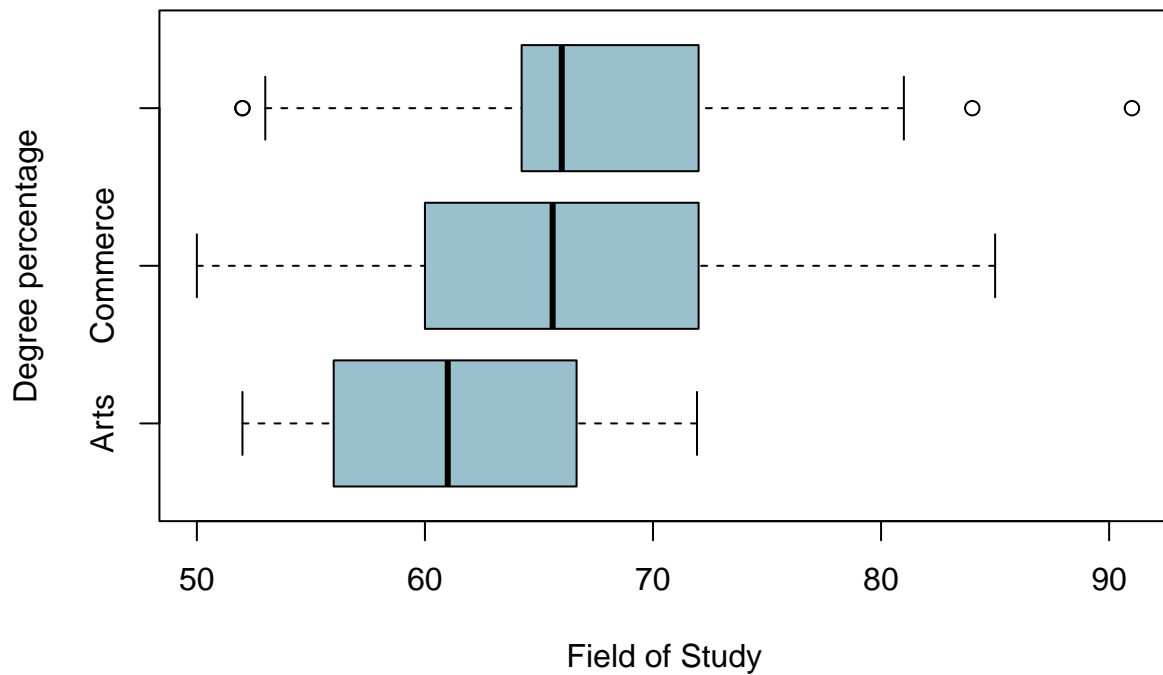
```
#Creates a boxplot with degree percentage.  
boxplot(phase2$degree_p, xlab="All Participants", ylab="Degree percentage",  
        vertical=TRUE, ylim=c(0,100), col="lightblue3")
```



Multivariate

*#Creates a boxplot with degree percentage and field of specialization in higher
#secondary education variables.*

```
boxplot(phase2$degree_p~phase2$hsc_s, ylab="Degree percentage",  
        xlab="Field of Study",horizontal=TRUE, col="lightblue3"))
```

What is the difference in MBA percentage between males and females who were placed?

Stripcharts vary in whether the variables are bivariate or multivariate.

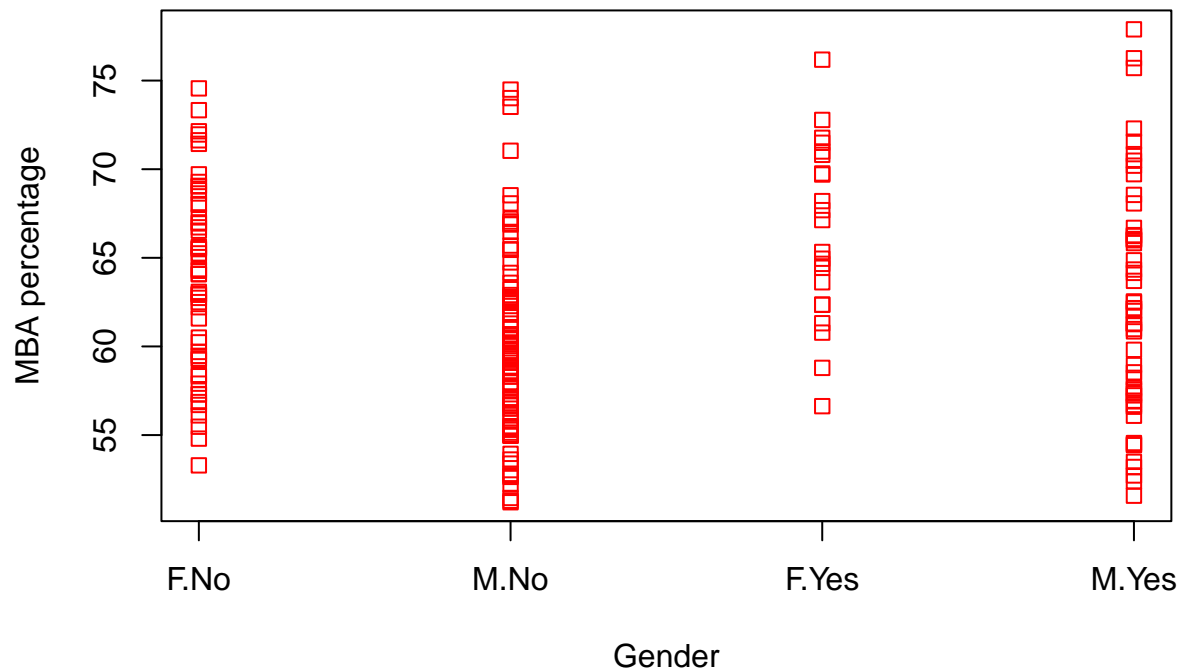
Bivariate

```
#Creates a stripchart with mba percentage and gender.
stripchart(phase2$mba_p~phase2$gender,ylab="MBA percentage", vertical=TRUE,
           xlab="Gender",pch=(0), col=("black"))
```



Multivariate

```
#Creates a stripchart with mba percentage,gender and work experience.
stripchart(phase2$mba_p~phase2$gender*phase2$workex, ylab="MBA percentage",
           vertical=TRUE,xlab="Gender", pch=(0),col=("red"))
```

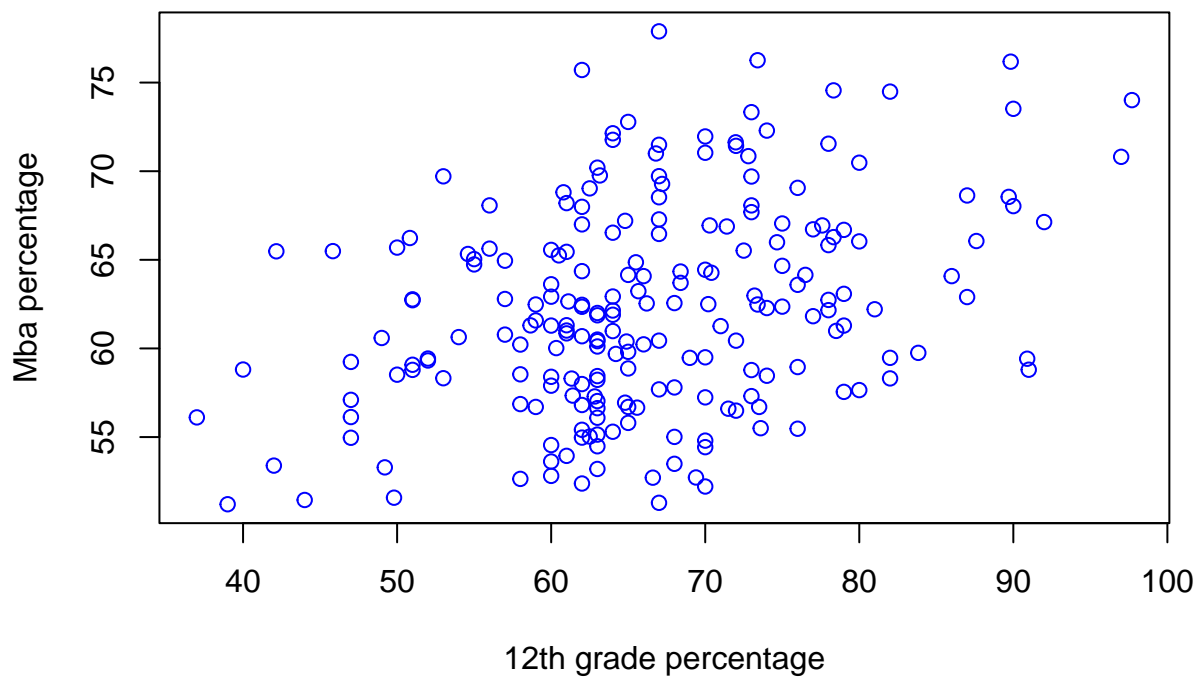


How does the 12th grade education percentage affect the MBA percentage?

Scatterplots vary in whether a categorical variable is added.

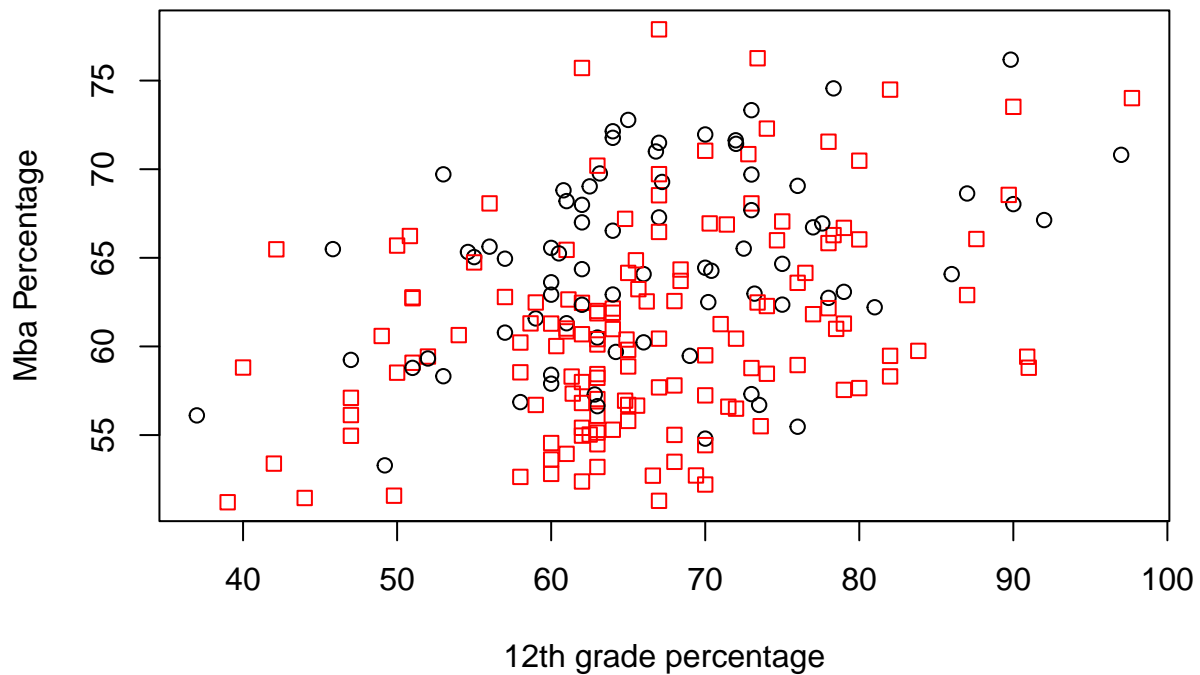
Bivariate

```
#Creates a scatterplot with higher secondary education percentage and mba
#percentage.
plot(phase2$hsc_p, phase2$mba_p, xlab="12th grade percentage",
     ylab="Mba percentage", col="blue")
```



Multivariate

```
#Creates a scatterplot with higher secondary education percentage, mba  
#percentage and gender. (with ifelse condition)  
plot(phase2$hsc_p, phase2$mba_p, xlab="12th grade percentage", ylab="Mba Percentage",  
      col=ifelse(phase2$gender=="M", "red", "black"),  
      pch=ifelse(phase2$gender=="M", 0, 1))
```



ANALYSIS SECTION

CONCLUSION SECTION

HELPFUL HINTS AND REMINDERS

1. Deactivate any antiviruses or bitdefenders prior to working with R, it might result in problems when trying to knit your PDF if not deactivated.
2. Ensure you have all appropriate packages before working on the assignment given.
3. Ensure the csv file is in a user file (Mac) that can be retrieved when importing your dataset into R.

VARIABLE NAMES AND DESCRIPTIONS

```
#Creates a matrix to store the table.
#nrow is the number of the rows.
StatsProject <- matrix(c("", "", "", "gender", "The gender of the students",
                          "Categorical, nominal",
                          "", "participating in the study", "",
```

```

"", "- Male=â€Mâ€,Female='F' ", "",
"", "", "",
"ssc_p", "The secondary education percentage",
"Quantitative, ratio ",
"", "received by participating students", "",
"", "when they were in 10th grade", "",
"", "Units:%", "",
"", "", "",
"ssc_b", "The board of education students",
"Categorical, nominal",
"", "attended in secondary school.Students", " ",
"", "who attended the Central board", "",
"", "= Central, students who did not", "",
"", "attend the central Board= Others", "",
"", "", "",
"hsc_p ",
"The higher secondary education",
"Quantitative, ratio",
"", "percentage received by participating", "",
"", "students when they were in 12th grade", "",
"", "Units:%", "",
"", "", "",
"hsc_b", "The board of education students",
"Categorical, nominal",
"", "attended in secondary school.Students", "",
"", "who attended the Central board", "",
"", "= Central, students who did not", "",
"", "attend the central Board= Others", "",
"", "", "",
"hsc_s", "The specialization of education of the",
"Categorical, nominal",
"", "students in secondary higher education", "",
"", "(12th grade)", "",
"", "", "",
"degree_p ", "The degree percentage of students", "Quantitative, ratio",
"", "graduating secondary school", "",
"", "Units:%", "",
"", "", "",
"degree_t",
"The degree type(feild of degree",
"Categorical, nominal",
"", "education)of the students in their", "",
"", "undergraduate", "",
"", "", "",
"workex", "A description of which students", "Categorical, nominal",
"", "have work experience (Yes/No)", "",
"", "", "",
"etest_p", "The employability test percentage of", "Quantitative, ratio",
"", "each student, conducted by the college", "",
"", "Units:%", "",
"", "", "",
"specialisation ", "Specialization of each student",
"Categorical, nominal",

```

```

    "", "post graduation (MBA)", "",
    "", "", "",
    "mba_p", "The degree percentage of", "Quantitative, ratio",
    "", "students graduating MBA", "",
    "", "Units:%", "",
    "", "", "",
    "status", "Status of placement in a job post",
    "Categorical, nominal",
    "", "MBA graduation- (Placed/Not placed)", "",
    "", "", "",
    "salary", "Salary offered by corporate ", "Quantitative, ratio",
    "", "to candidates", "",
    "", "Unit: Rupee", ""), nrow = 59, byrow=TRUE)
#Column names of the matrix.
colnames(StatsProject) <- c("VARIABLE", "DESCRIPTION", "TYPE")
#Row names of the matrix.
rownames(StatsProject) <- c("", "1", "", "", "", "2", "", "", "", "", "3", "", "", "", "", "", "4",
    "", "", "", "", "5", "", "", "", "", "", "6",
    "", "", "", "7", "", "", "", "8", "", "", "", "9", "", "", "10", "", "", "", "11",
    "", "", "12", "", "", "", "13", "", "", "14", "", "")
#Creates a table with matrix.
StatsProject <- as.table(StatsProject)
StatsProject

```

##	VARIABLE	DESCRIPTION	TYPE
##			
## 1	gender	The gender of the students participating in the study	Categorical, nominal
##		- Male=â\200\231Mâ\200\231,Female='F'	
##			
## 2	ssc_p	The secondary education percentage received by participating students when they were in 10th grade	Quantitative, ratio
##		Units:%	
##			
## 3	ssc_b	The board of education students attended in secondary school.Students who attended the Central board	Categorical, nominal
##		= Central, students who did not attend the central Board= Others	
##			
## 4	hsc_p	The higher secondary education percentage received by participating students when they were in 12th grade	Quantitative, ratio
##		Units:%	
##			
## 5	hsc_b	The board of education students attended in secondary school.Students who attended the Central board	Categorical, nominal
##		= Central, students who did not attend the central Board= Others	
##			
## 6	hsc_s	The specialization of education of the	Categorical, nominal

##		students in secondary higher education	
##		(12th grade)	
##			
## 7	degree_p	The degree percentage of students	Quantitative, ratio
##		graduating secondary school	
##		Units: %	
##			
## 8	degree_t	The degree type (field of degree	Categorical, nominal
##		education) of the students in their	
##		undergraduate	
##			
## 9	workex	A description of which students	Categorical, nominal
##		have work experience (Yes/No)	
##			
## 10	etest_p	The employability test percentage of	Quantitative, ratio
##		each student, conducted by the college	
##		Units: %	
##			
## 11	specialisation	Specialization of each student	Categorical, nominal
##		post graduation (MBA)	
##			
## 12	mba_p	The degree percentage of	Quantitative, ratio
##		students graduating MBA	
##		Units: %	
##			
## 13	status	Status of placement in a job post	Categorical, nominal
##		MBA graduation- (Placed/Not placed)	
##			
## 14	salary	Salary offered by corporate	Quantitative, ratio
##		to candidates	
##		Unit: Rupee	

REFERENCE

Roshan, B. (2019). Campus Recruitment: Academic and Employability Factors influencing placement. Kaggle. Retrieved October Friday, 15, from <https://www.kaggle.com/benroshan/factors-affecting-campus-placement>