Campus Recruitment

Academic and Employability Factors Influencing Placements

Kapil Rana

19/10/2021

```
library(tidyverse)
```

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

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

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

```
library(dlookr)
```

```
## Imported Liberation Sans Narrow fonts.
```

```
##
## Attaching package: 'dlookr'
```

```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
## The following object is masked from 'package:base':
##
##     transform
```

```
library(ggthemes)
library(rstatix)
```

```
##
## Attaching package: 'rstatix'
```

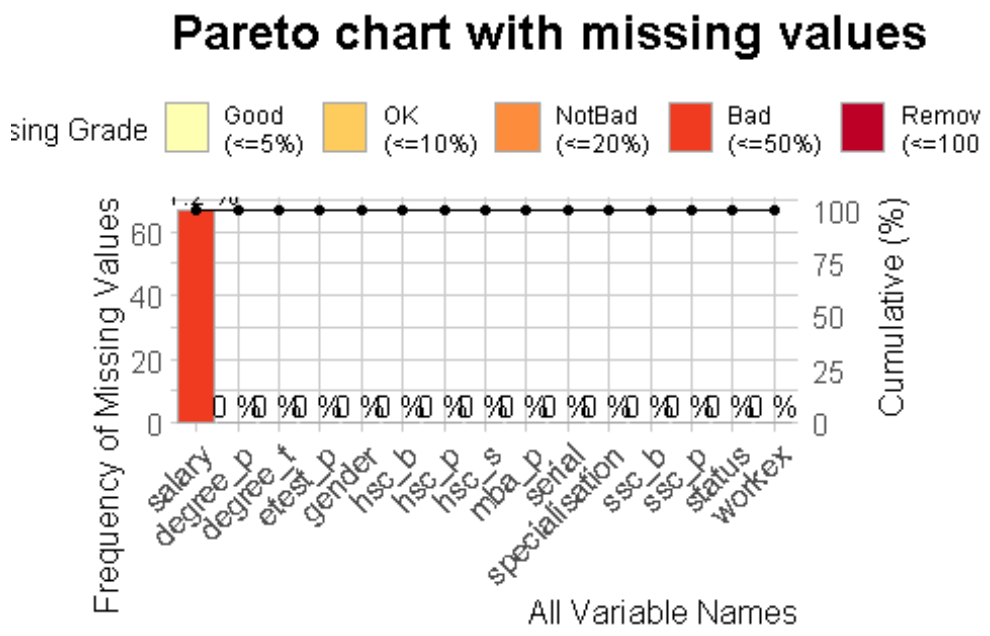
```
## The following object is masked from 'package:stats':
##
##     filter
```

```
setwd("C:\\Users\\kapil\\Desktop\\(Project)\\R-practice\\Placement Data")
getwd()
```

```
## [1] "C:/Users/kapil/Desktop/(Project)/R-practice/Placement Data"

placement <- readr::read_csv("Placement.csv")

##
## -- Column specification
--------------------------------------------------------
## cols(
##   sl_no = col_double(),
##   gender = col_character(),
##   ssc_p = col_double(),
##   ssc_b = col_character(),
##   hsc_p = col_double(),
##   hsc_b = col_character(),
##   hsc_s = col_character(),
##   degree_p = col_double(),
##   degree_t = col_character(),
##   workex = col_character(),
##   etest_p = col_double(),
##   specialisation = col_character(),
##   mba_p = col_double(),
##   status = col_character(),
##   salary = col_double()
## )

View(placement)

placement <- placement %>%
  rename("serial" = sl_no)

#About Data
columns <- names(placement)

description <- c(
"Serial",
"Gender: Male='M', Female='F'",
"Secondary Education Percentage (grades 9 and 10) - exam at the end of 10th
grade",
"Board of Education - Central/Others",
"Higher Secondary Education (grades 11 and 12) - exam at the end of 12th
grade",
"Board of Education - Central/Others",
"Specialization in Higher Secondary Education",
"Degree Percentage",
"Undergraduate (Degree type) - Field of degree education",
"Work Experience - Yes/No",
"Employability test Percentage (conducted by college)",
"Post Graduation (MBA) - Specialization",
"MBA percentage",
"Status of placement - Placed/Not placed",
```

```
"Salary offered by corporate to candidates")

dataset_table <-as.data.frame(cbind(columns,  description))
dataset_table

##           columns
## 1          serial
## 2          gender
## 3           ssc_p
## 4           ssc_b
## 5           hsc_p
## 6           hsc_b
## 7           hsc_s
## 8        degree_p
## 9        degree_t
## 10         workex
## 11        etest_p
## 12 specialisation
## 13          mba_p
## 14         status
## 15         salary
##
description
## 1
Serial
## 2                                                          Gender: Male='M',
Female='F'
## 3  Secondary Education Percentage (grades 9 and 10) - exam at the end of
10th grade
## 4                                                 Board of Education -
Central/Others
## 5     Higher Secondary Education (grades 11 and 12) - exam at the end of
12th grade
## 6                                                 Board of Education -
Central/Others
## 7                                    Specialization in Higher Secondary
Education
## 8                                                                Degree
Percentage
## 9                          Undergraduate (Degree type) - Field of degree
education
## 10                                                          Work Experience
- Yes/No
## 11                            Employability test Percentage (conducted by
college)
## 12                                        Post Graduation (MBA) -
Specialization
## 13                                                               MBA
percentage
## 14                                       Status of placement -
```

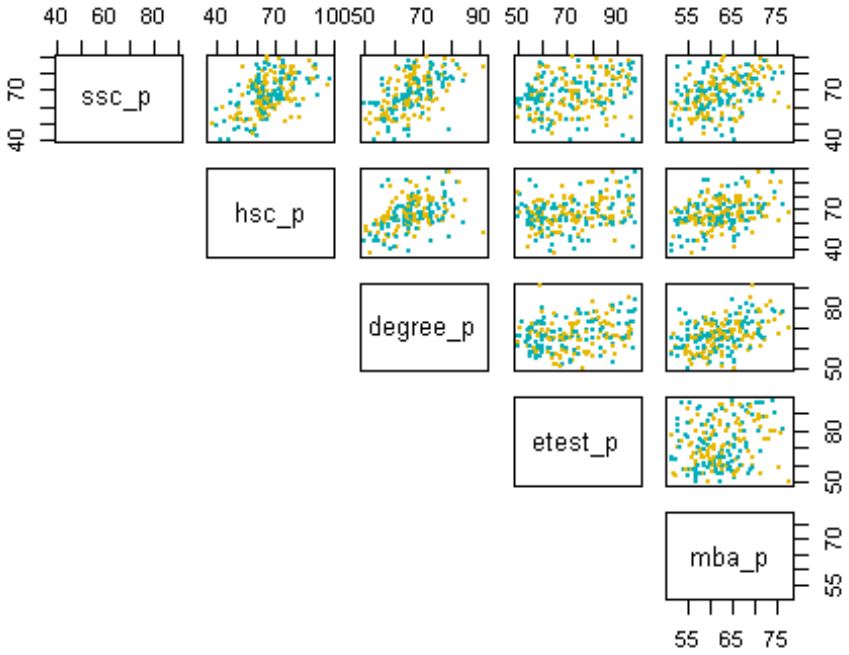
```
Placed/Not placed
## 15                                        Salary offered by corporate to
candidates

#Check for NA values
sapply(placement, function(x) sum(is.na(x)))

##         serial         gender          ssc_p          ssc_b          hsc_p

##              0              0              0              0              0

##          hsc_b          hsc_s       degree_p       degree_t         workex

##              0              0              0              0              0

##        etest_p specialisation          mba_p         status         salary

##              0              0              0              0             67

#Only salary column has 67 NA values
dlookr::plot_na_pareto(placement)
```




```
#As 67 students are not placed so we have 67 NA's in Salary Column
placement %>%
  count(status, sort = T)
```

```
## # A tibble: 2 x 2
##   status         n
##   <chr>      <int>
## 1 Placed       148
## 2 Not Placed    67
```

```
#Correlation between numerical variables
data_cor <- placement %>%
  select(ssc_p, hsc_p, degree_p, etest_p, mba_p)

my_cols <- c("#00AFBB", "#E7B800")
pairs(data_cor, pch = 19,  cex = 0.5,
      col = my_cols,
      lower.panel=NULL)
```

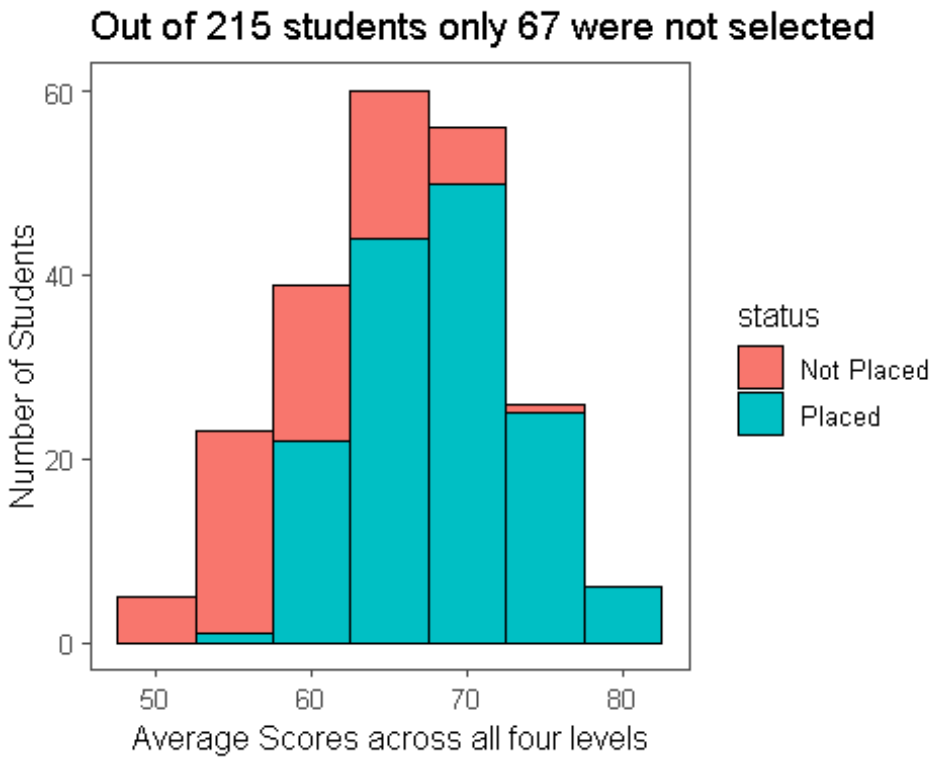



```
#There is a medium correlation between academic scores it means those who performed well in their secondary education also perform well with higher secondary and bachelor degree but employability test scores only had a low correlation with academic scores.
data <- as.data.frame(cor(data_cor))
data
```

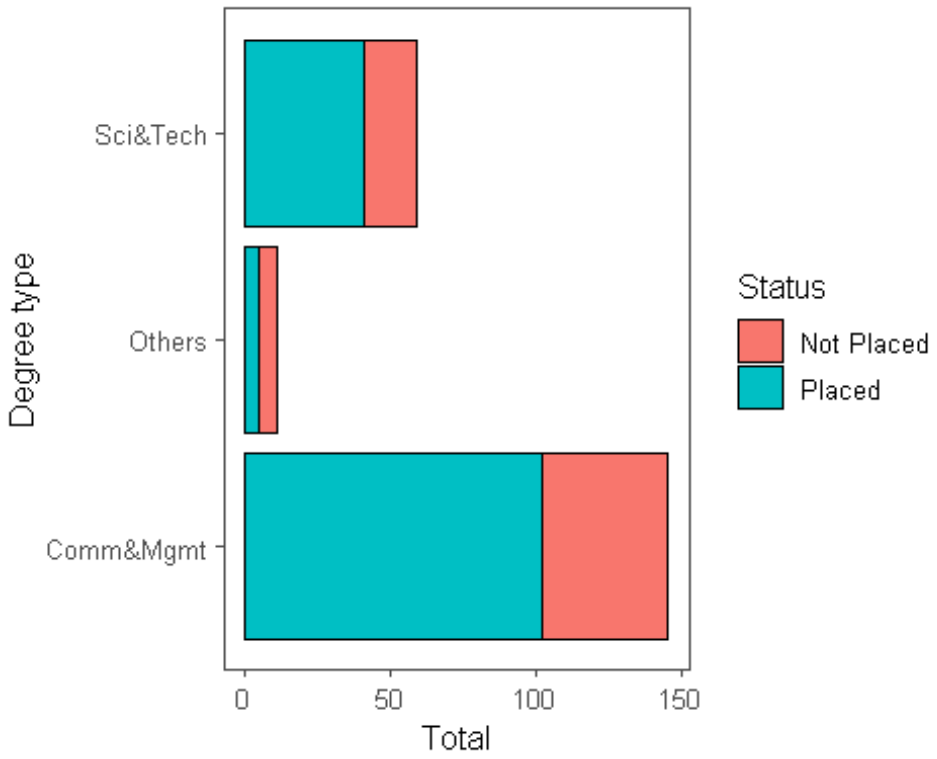
```
##              ssc_p     hsc_p  degree_p   etest_p     mba_p
## ssc_p    1.0000000 0.5114721 0.5384040 0.2619927 0.3884776
## hsc_p    0.5114721 1.0000000 0.4342058 0.2451129 0.3548226
## degree_p 0.5384040 0.4342058 1.0000000 0.2244702 0.4023638
```

```
## etest_p  0.2619927 0.2451129 0.2244702 1.0000000 0.2180547
## mba_p    0.3884776 0.3548226 0.4023638 0.2180547 1.0000000
```

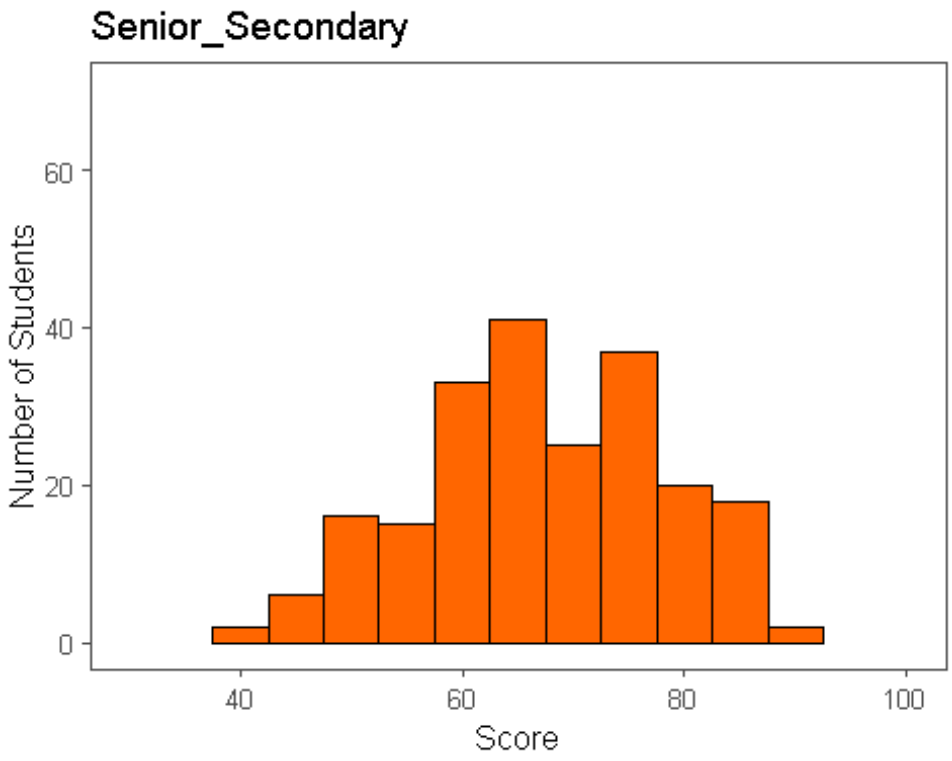
```
#Check the Percentage of the students with Status
#Out of 215 Students only 148 students were  placed
placement %>%
  mutate(total = (ssc_p + hsc_p + degree_p + mba_p)/4) %>%
  ggplot(aes(total, fill = status)) +
  geom_histogram(binwidth = 5, col = "black") +
  labs(x = "Average Scores across all four levels",
       y = "Number of Students",
       title = "Out of 215 students only 67 were not selected")+
  ggthemes::theme_few()
```



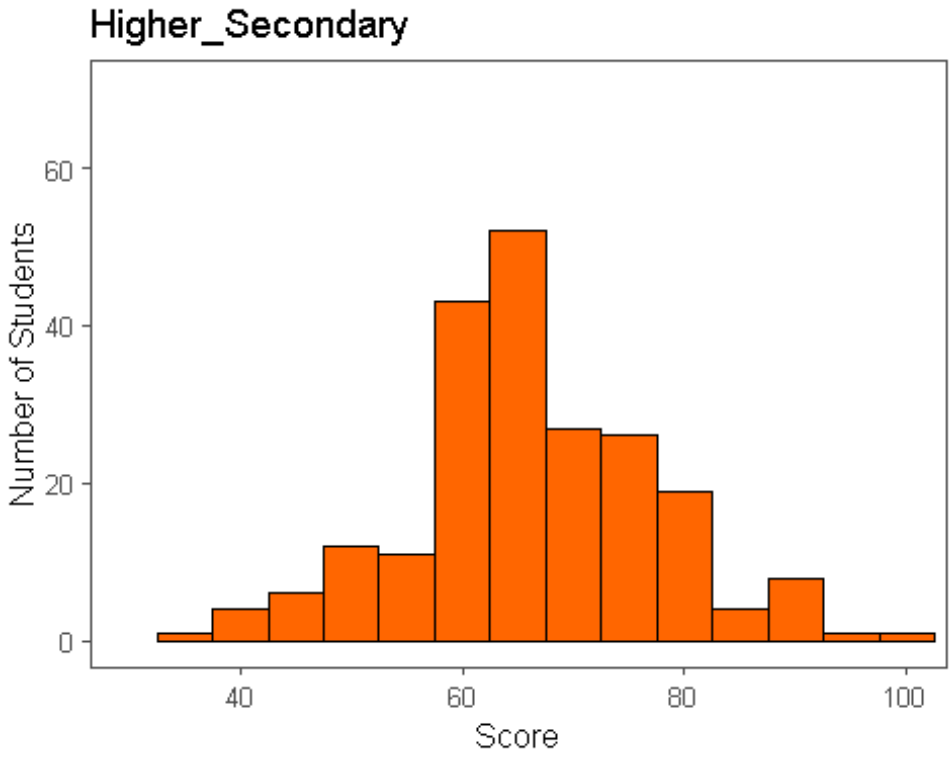

```
#Most no. of student who get placed are from commerce background
placement %>%
  ggplot(aes(degree_t)) +
  geom_bar(aes(fill = placement$status), col = "black") +
  coord_flip() +
  labs(x = "Degree type",
       y = "Total",
       fill = guide_legend(title = "Status")) +
  ggthemes::theme_few()
```



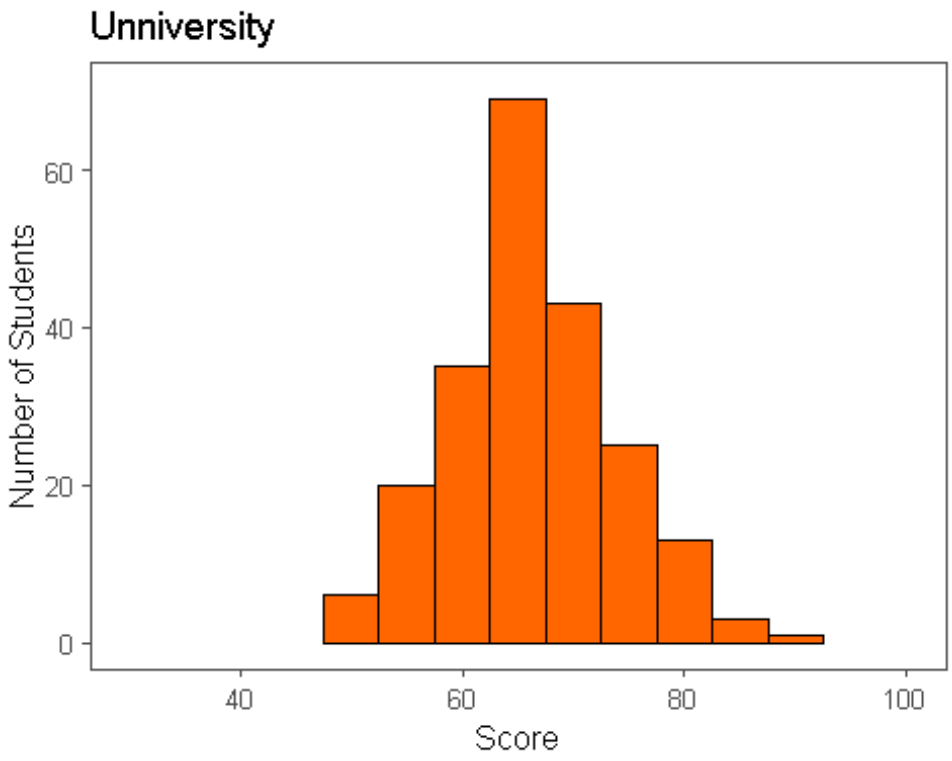

```
#Lets check the distribution of Scores look like for each level of education
Employability test has a different distribution than the others with very
wide and almost equal distribution of each bucket

distribution <- function(level, Lev_name) {
  placement %>%
    ggplot(aes(level)) +
    geom_histogram(binwidth = 5, fill = "#ff6600" ,col = "black") +
    coord_cartesian(xlim = c(30,100),
                    ylim = c(0,70)) +
    labs(x = "Score",
         y = "Number of Students",
         title = Lev_name) +
    ggthemes::theme_few()
}


distribution(placement$ssc_p, "Senior_Secondary")
```



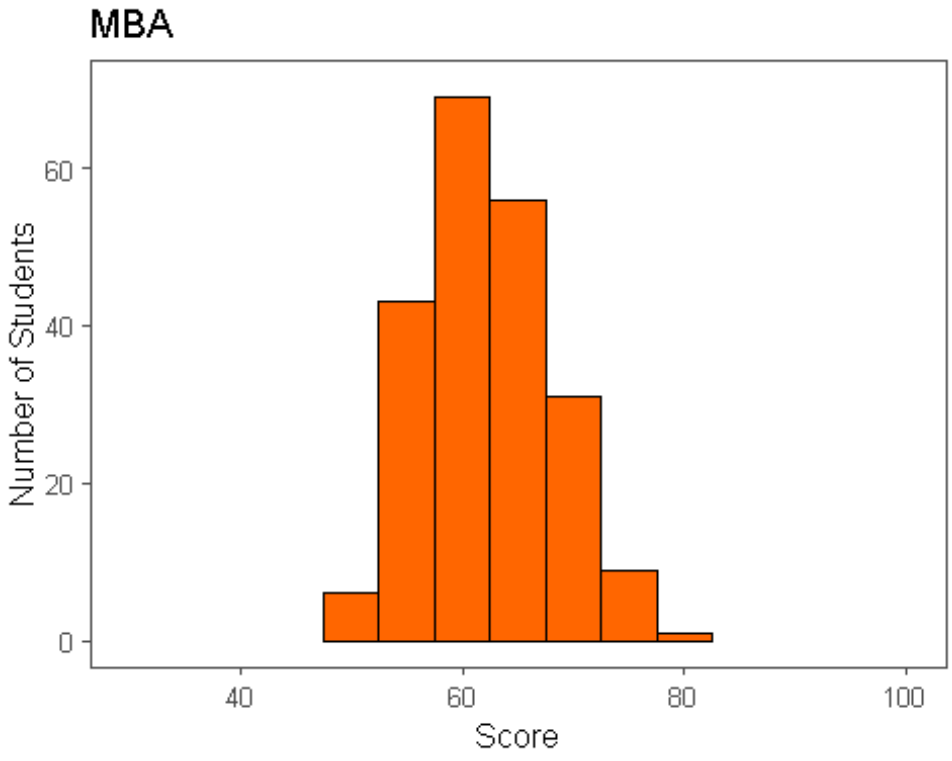

```
distribution(placement$hsc_p, "Higher_Secondary")
```



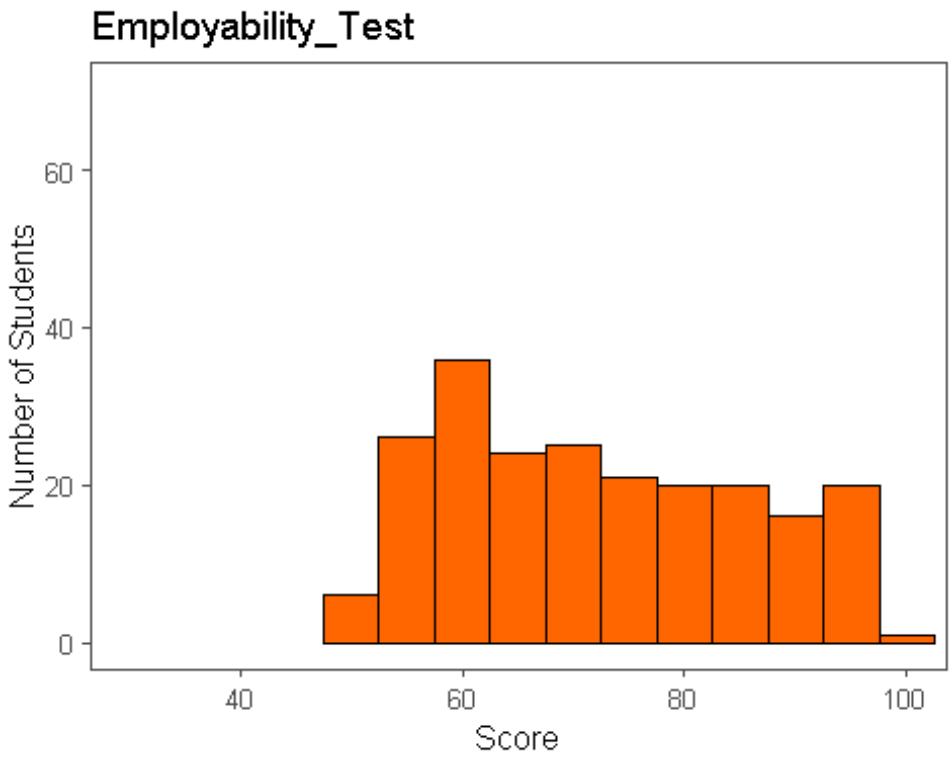

```
distribution(placement$degree_p, "Unniversity")
```



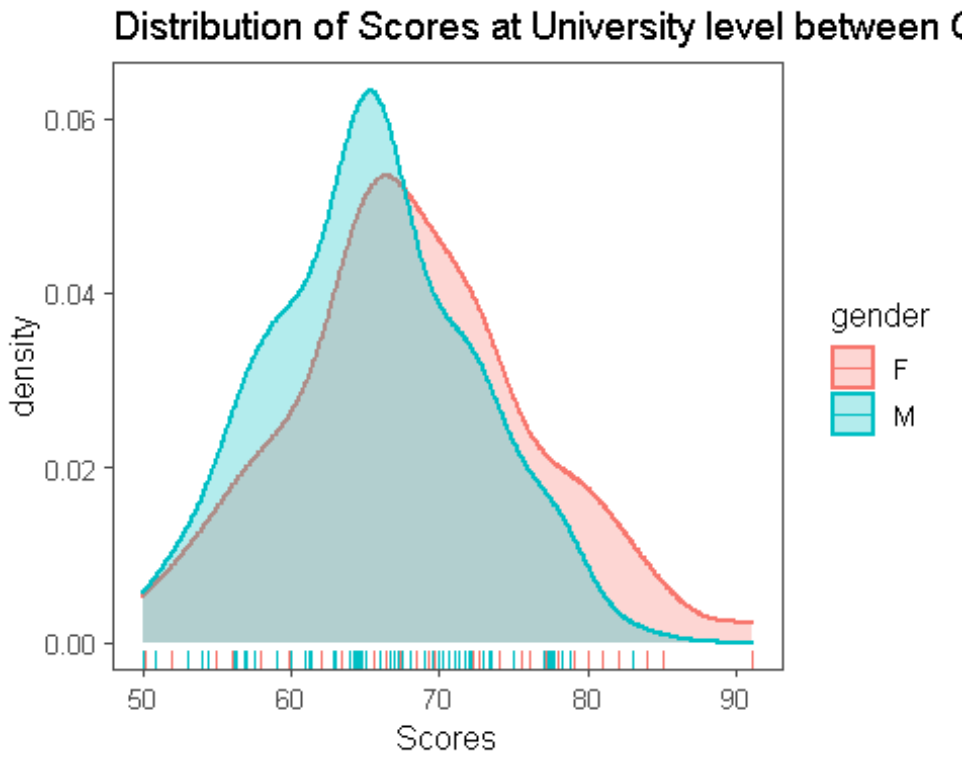

```
distribution(placement$mba_p, "MBA")
```



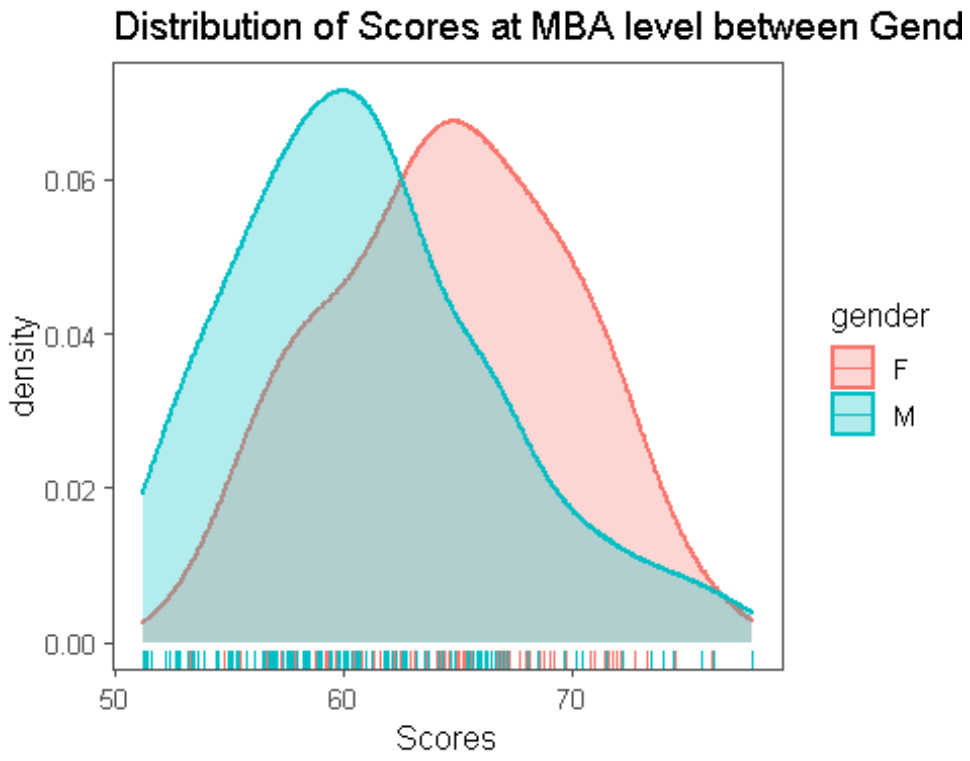

```
distribution(placement$etest_p, "Employability_Test")
```



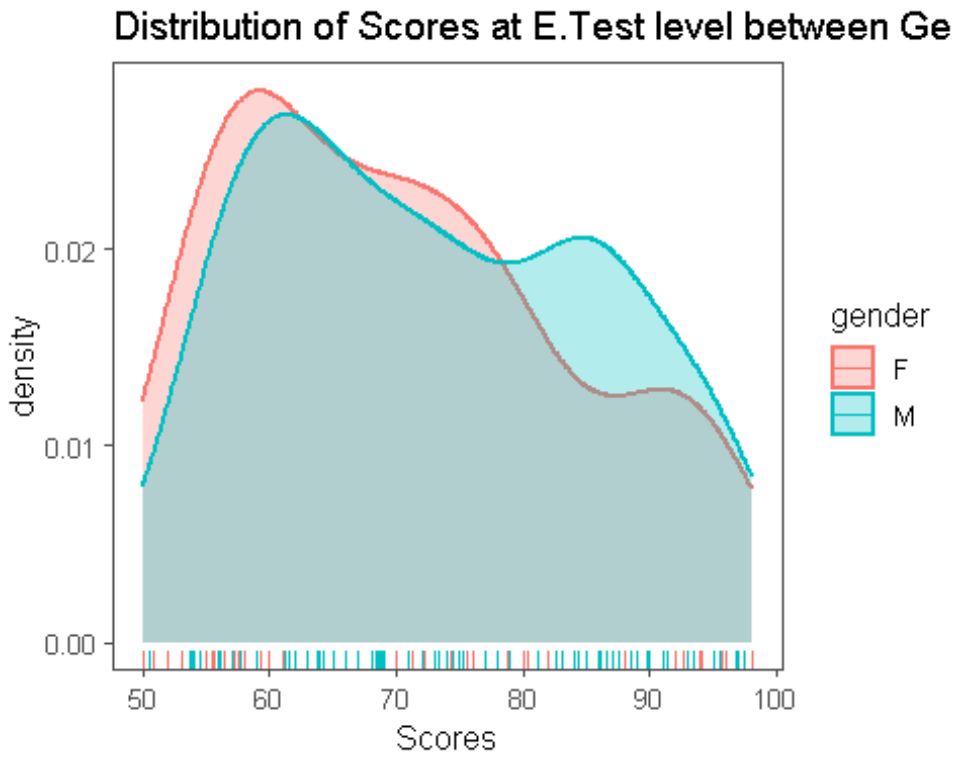

```
#Is there any significant difference between Genders and their Academic &
E.test score?
#Density plot shows that females score higher than the male but the t.test
gave the p.value lower than the alpha(0.05) We have to reject the null
hypothesis,It means there is no significant differences in performance
##during University and MBA levels

#Lets Find Out

placement %>%
  ggplot(aes(degree_p, fill = gender, col = gender)) +
  geom_density(alpha = 0.3, lwd = 1, show.legend = T) +
  geom_rug() +
  labs(x = "Scores",
       title = "Distribution of Scores at University level between Genders")
+
  ggthemes::theme_few()
```




```
placement %>%
  ggplot(aes(mba_p, fill = gender, col = gender)) +
  geom_density(alpha = 0.3, lwd = 1, show.legend = T) +
  geom_rug() +
  labs(x = "Scores",
       title = "Distribution of Scores at MBA level between Genders") +
  ggthemes::theme_few()
```




```
placement %>%
  ggplot(aes(etest_p, fill = gender, col = gender)) +
  geom_density(alpha = 0.3, lwd = 1, show.legend = T) +
  geom_rug() +
  labs(x = "Scores",
       title = "Distribution of Scores at E.Test level between Genders") +
  ggthemes::theme_few()
```




```
##Now check the significance with a statistical test
placement %>%
  t_test(degree_p ~ gender, conf.level = 0.95) %>%
  add_significance()
```

```
## # A tibble: 1 x 9
##   .y.      group1 group2    n1    n2 statistic    df      p p.signif
## * <chr>    <chr>  <chr>  <int> <int>     <dbl> <dbl>  <dbl> <chr>
## 1 degree_p F      M         76   139      2.43  132. 0.0164 *
```

```
placement %>%
  t_test(mba_p ~ gender, conf.level = 0.95)
```

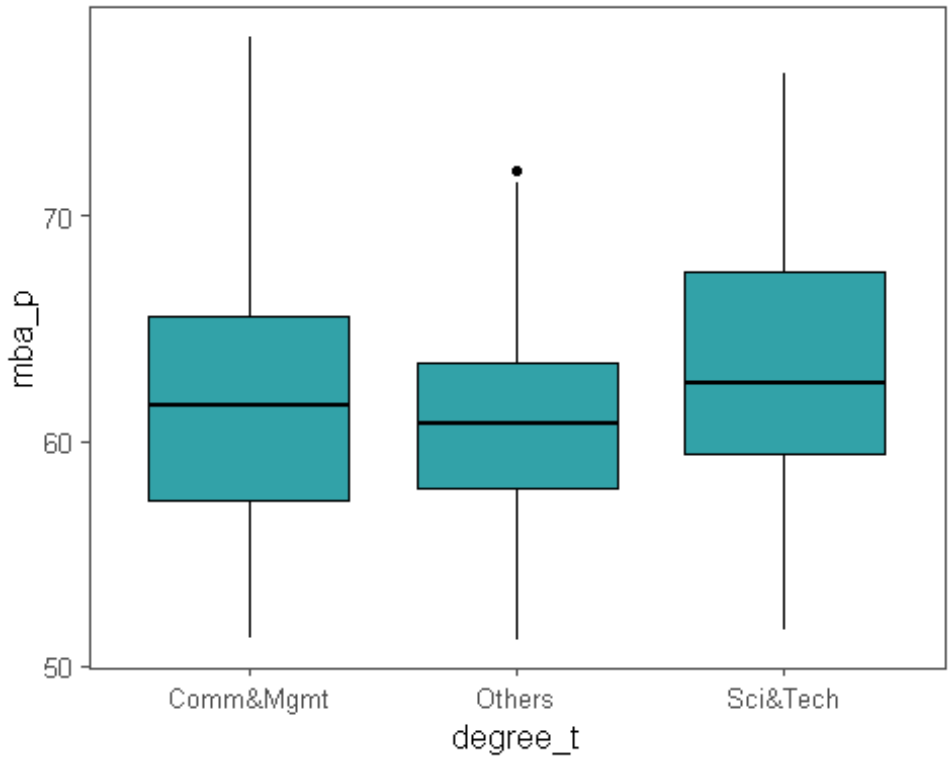
```
## # A tibble: 1 x 8
##   .y.   group1 group2    n1    n2 statistic    df          p
## * <chr> <chr>  <chr>  <int> <int>     <dbl> <dbl>      <dbl>
## 1 mba_p F      M         76   139      4.73  167. 0.00000486
```

```
#As the score became more concentrated around median as the student
progressed in their education, we could infer that there was less chance of
differentiating themselves based on grade and more based on other factors
such as : employability score, work experience, technical skills, soft
skills, better interviewing skills, extra-curricular experience

#ANOVA  test to check whether Type of Degree(commerce, science & arts) had
any significant impact on the avg. MBA marks
```

```r
#As p.value is more than our critical value 0.05, we can conclude that mean
of MBA marks under dufferent degree types are same
aov(mba_p ~ degree_t, data = placement) %>%
  summary()
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## degree_t      2    116   58.13    1.72  0.182
## Residuals   212   7166   33.80
```

```r
#Visualize
placement %>%
  ggplot(aes(degree_t, mba_p)) +
  geom_boxplot(fill = "#32a2a8", col = "black") +
  ggthemes::theme_few()
```




```r
#Bartlett's  test to check homegeniety of variance among Types of Degrees
with regarding to MBA marks
#As p.value is more than our critical value 0.05, it suggest that the
variance among groups is equal
bartlett.test(as.numeric(mba_p) ~ as.factor(degree_t), data = placement)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  as.numeric(mba_p) by as.factor(degree_t)
## Bartlett's K-squared = 0.15907, df = 2, p-value = 0.9235
```

```r
#ANOVA  test to check whether Type of Degree(commerce, science & arts) had any significant impact on the Salary

#As p value is more than 0.05,  We can conclude that the mean salary under different degree type are same
aov(salary ~ degree_t, data = placement) %>%
  summary()
```

```
##              Df    Sum Sq   Mean Sq F value Pr(>F)
## degree_t      2 3.822e+10 1.911e+10   2.224  0.112
## Residuals   145 1.246e+12 8.591e+09
## 67 observations deleted due to missingness
```

```r
#Bartlett's  test to check homegeniety of variance among Types of Degrees with regarding Salary
#As p.value is more than our critical value 0.05, it suggest that the variance among groups is equal
bartlett.test(as.numeric(salary) ~ as.factor(degree_t), data = placement)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  as.numeric(salary) by as.factor(degree_t)
## Bartlett's K-squared = 2.5268, df = 2, p-value = 0.2827
```

```r
#Chi-square test to check whether the type of degree has any impact on Status(Placed or not placed), we are using chi2 test as both of our column are factor column
contigency <- with(placement, table(placement$degree_t, placement$status))
contigency <- as.matrix(contigency)
contigency
```

```
##
##             Not Placed Placed
##   Comm&Mgmt         43    102
##   Others             6      5
##   Sci&Tech          18     41
```

```r
#As p.value is more than our critical value 0.05, we conclude that two categorical variables are independent in some population.
chi.test <- chisq.test(contigency, correct = F)
chi.test
```

```
##
##  Pearson's Chi-squared test
##
## data:  contigency
## X-squared = 2.969, df = 2, p-value = 0.2266
```