

**R-Project**

**Master of Business Administration BA**

**Batch 2023-25**

**SUBMITTED TO: SUBMITTED BY:**

MAHENDRA SINGH YADAV VISHAKHA JEEVTANI

PIMRG

##### **DECLARATION**

I  **VISHAKHA JEEVTANI** student of MBA BA Sem.2 Section A of Prestige Institute of Management & Research, Gwalior, hereby declare that the R PROJECT entitled **“Campus Recruitement ”** is submitted by me in the line of partial fulfillment of course objectives for the Masters of Business Administration BA.

I assure that this report is the result of our efforts and that any other institute for awarding any degree or diploma has not submitted it.

Date: 19/04/2024 Name : MAHENDRA SINGH YADAV

Place: PIMRG

#### CERTIFICATE

# This is to certify that, Vishakha Jeevtani of MBA BA Sem 2 Section A of Prestige Institute of Management & Research Gwalior, have successfully completed their report. They have prepared this report entitled **“****Campus Recruitment”** under my direct supervision and guidance.

Date: 19/04/24

Place: PIMRG

**ACKNOWLEDGEMENT**

I, **VISHAKHA JEEVTANI** express my sincere gratitude to MAHENDRA SINGH YADAV giving us the opportunity to work under his guidance on the project entitled **“Campus Recruitment”.**

I am grateful to our Director Prof. Nishant Joshi, Deputy Director Prof. Tarika Singh, Faculty Members and our friends for their valuable suggestions in the execution of project preparation.

I am also thankful to all the faculty members, friends and other staff that guided and helped us very kindly at each and every step whenever we required.

I also acknowledge & convey thanks to the library staff, and computer department of PIMRG for their kind and valuable support.

**“VISHAKHA JEEVTANI”**

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S No** | **Content** | **Page No** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**SECTION - A**

1. What do you mean by Exploratory Data Analytics? Explain

Exploratory Data Analysis (EDA) is an approach to analyzing datasets to summarize their main characteristics, often with visual methods. It is a crucial step in the data analysis process, especially in the initial stages of understanding the data before applying more advanced techniques.

The main objectives of exploratory data analysis are:

1. **Data Familiarization**: EDA helps analysts become familiar with the dataset they are working with. This involves understanding the variables, their types, distributions, and potential relationships.
2. **Detecting Patterns and Anomalies**: EDA involves identifying patterns and trends in the data, as well as any outliers or anomalies that may require further investigation.
3. **Formulating Hypotheses**: By exploring the data visually and statistically, analysts can form hypotheses about relationships between variables or potential factors that may influence certain outcomes.
4. **Selecting Appropriate Techniques**: EDA helps in deciding which statistical techniques or machine learning models are suitable for further analysis. Understanding the data's characteristics can guide the selection of appropriate methodologies.
5. **Data Cleaning and Preprocessing**: During EDA, analysts often identify missing values, inconsistencies, or errors in the data, which can then be addressed through data cleaning and preprocessing steps.
6. **Communicating Insights**: EDA produces visualizations and summaries that can effectively communicate insights to stakeholders, such as decision-makers or other members of the team.

EDA techniques include summary statistics, data visualization (such as histograms, box plots, scatter plots, and heatmaps), correlation analysis, dimensionality reduction techniques (like PCA), and clustering methods. These techniques help analysts gain insights into the underlying structure of the data, its distribution, relationships between variables, and potential patterns or trends. Overall, EDA serves as a crucial foundation for more advanced analytics and decision-making processes.

Top of Form

2. Describe “Test of Hypothesis”?

A Test of Hypothesis, also known as a hypothesis test or statistical test, is a method used in statistics to make decisions about the validity of a hypothesis or claim about a population parameter. In other words, it helps determine whether there is enough evidence in the sample data to support or reject a specific assertion about a population characteristic.

The process of hypothesis testing typically involves the following steps:

1. **Formulating Hypotheses**:
   * Null Hypothesis (H0): This is the default hypothesis, often denoted as H0, which states that there is no effect, no difference, or no relationship in the population. It represents the status quo or the absence of an effect.
   * Alternative Hypothesis (Ha or H1): This is the assertion or claim that is being tested against the null hypothesis. It represents what the researcher is trying to find evidence for.
2. **Selecting a Significance Level (α)**:
   * The significance level, denoted by α, determines the probability of making a Type I error (rejecting the null hypothesis when it is actually true). Commonly used significance levels include 0.05 and 0.01.
3. **Choosing a Test Statistic**:
   * The test statistic is a numerical value calculated from the sample data that is used to assess the plausibility of the null hypothesis. The choice of test statistic depends on the nature of the data and the hypothesis being tested.
4. **Calculating the P-value**:
   * The P-value is the probability of obtaining a test statistic as extreme as, or more extreme than, the observed value under the assumption that the null hypothesis is true. It provides a measure of the strength of evidence against the null hypothesis.
5. **Making a Decision**:
   * If the P-value is less than or equal to the significance level (α), then there is sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis.
   * If the P-value is greater than the significance level, then there is not enough evidence to reject the null hypothesis.
6. **Interpreting the Results**:
   * Based on the decision made in the previous step, conclusions are drawn regarding the hypothesis being tested. If the null hypothesis is rejected, it suggests that there is evidence to support the alternative hypothesis.

Hypothesis testing is a fundamental tool in statistical inference and is widely used in various fields such as science, engineering, business, and social sciences to draw conclusions from sample data about populations.

**SECTION – B**

## **About Dataset**

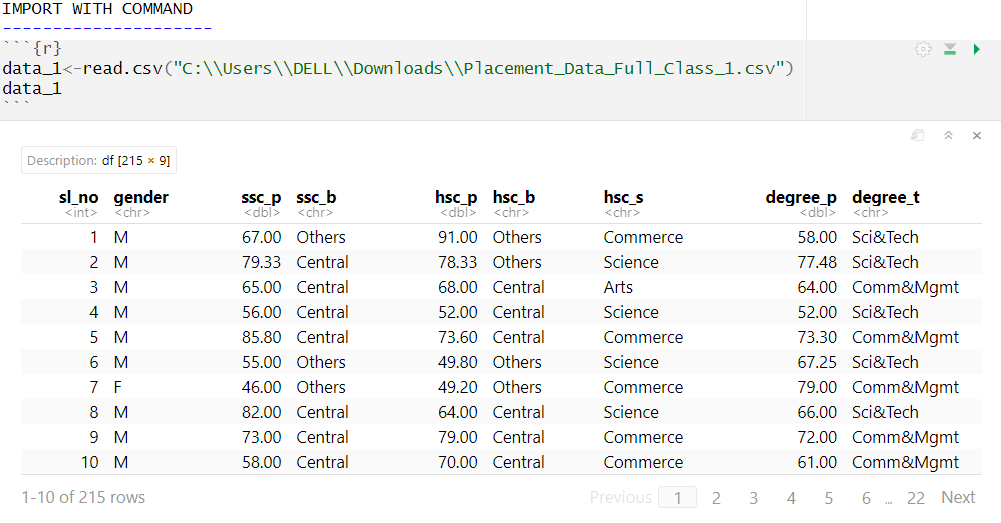
This data set consists of Placement data of students in a XYZ campus. It includes secondary and higher secondary school percentage and specialization. It also includes degree specialization, type and Work experience and salary offers to the placed students

**Attributes of Dataset**

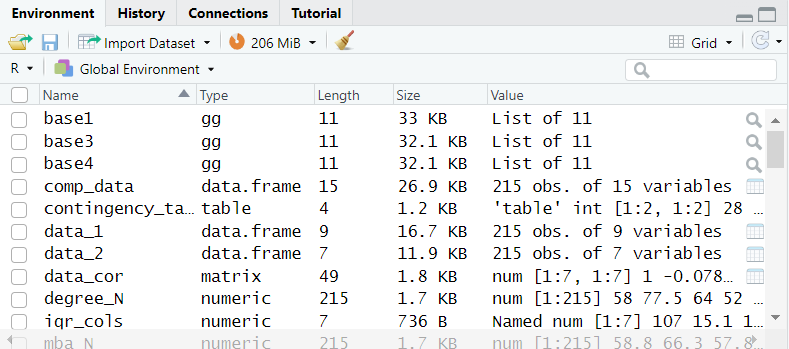
1. sl \_no – Serial Number
2. gender – Gender ‘MALE = M, FEMALE = F’
3. ssc\_p – Secondary Education percentage
4. ssc\_b – Board of secondary Education ‘Central/Others’
5. hsc\_p – Higher secondary Education percentage
6. hsc\_b – Board of Higher secondary Education ‘Central/Others’
7. hsc\_s – Specialization in Higher secondary Education ‘Science/Commerce/Arts’
8. degree\_p – Degree Percentage
9. degree\_t – Under graduation ‘Degree type - comm/mgmt, sci/tech, others’
10. workex – Work Experience
11. etest\_p – Employability test percentage
12. specialization - Post Graduate (MBA) Specialization ‘mkt&fin, mkt&hr’
13. mba\_p – MBA percentage
14. status – Status of placement ‘Placed/Not Placed’
15. salary – Salary offered by corporate to candidate

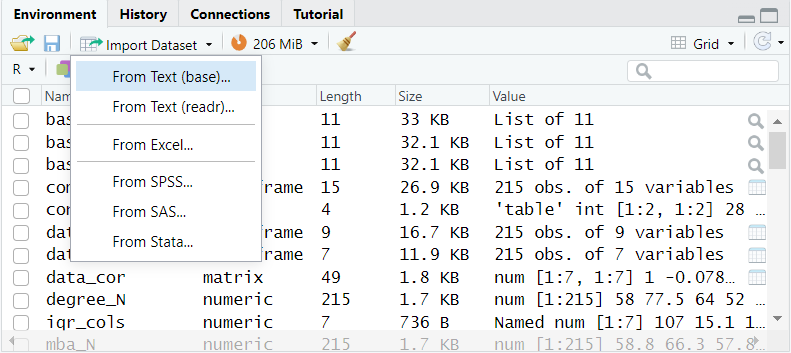
Additional, I divided the dataset into two same size dataset with connecting variable sl\_no to perform some merge operations.

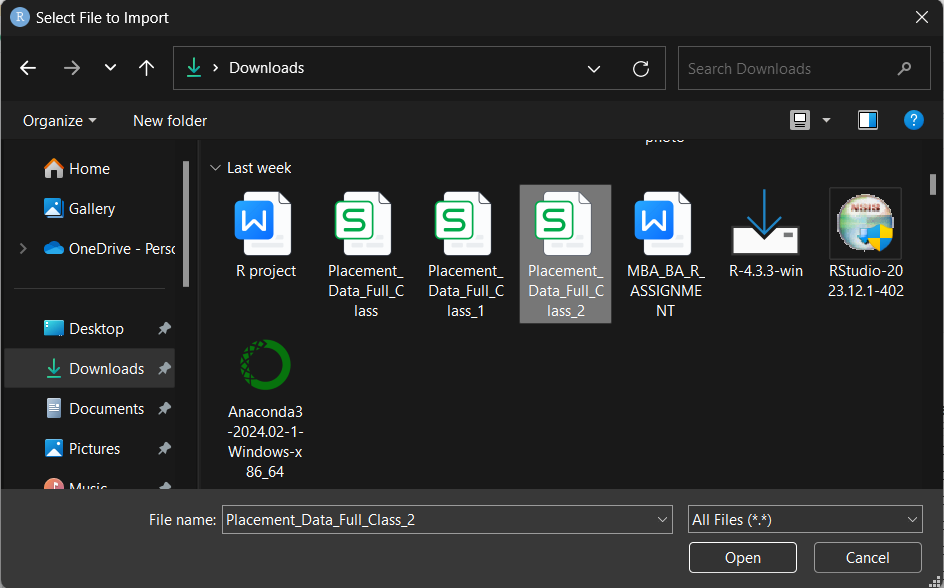
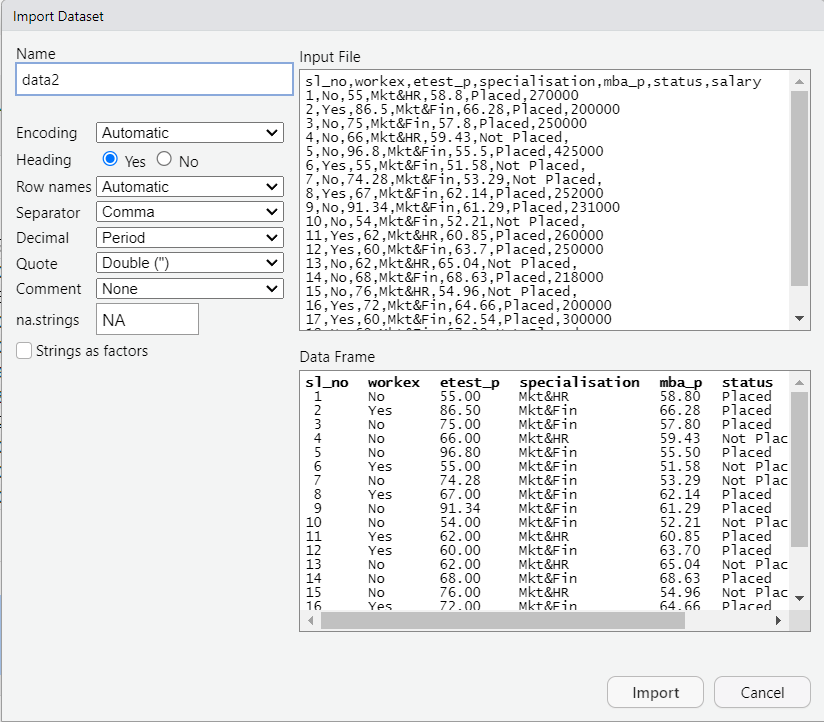
Import the dataset with command and without command

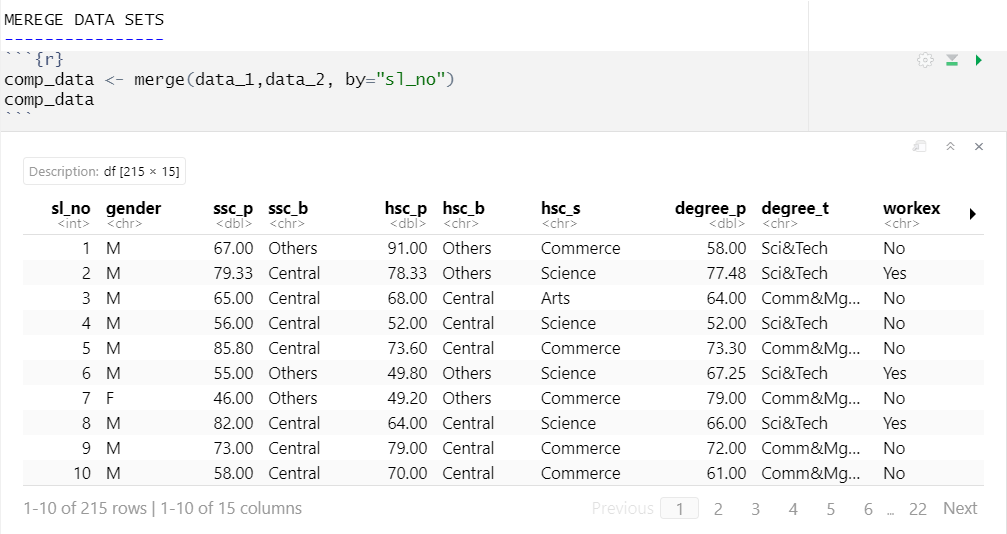


Without Command

* select Import dataset in Environment.
* Select From Text(base)

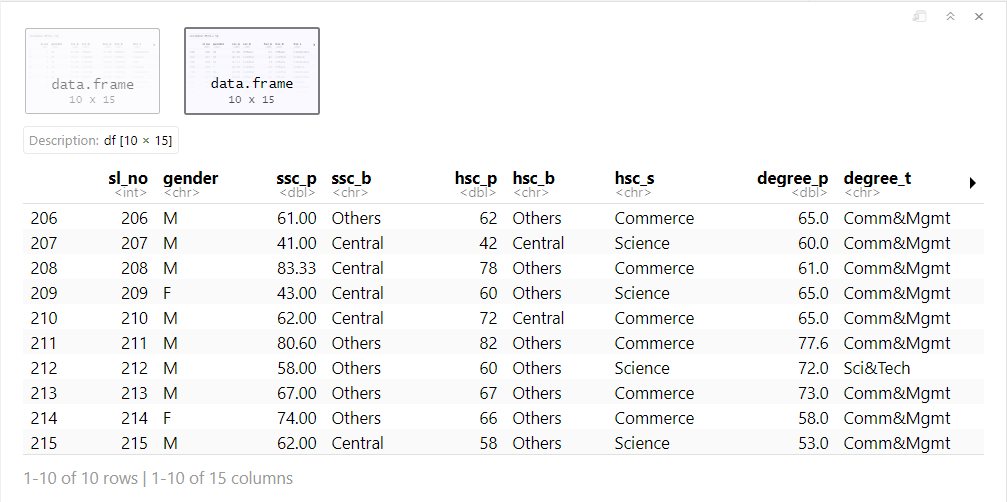
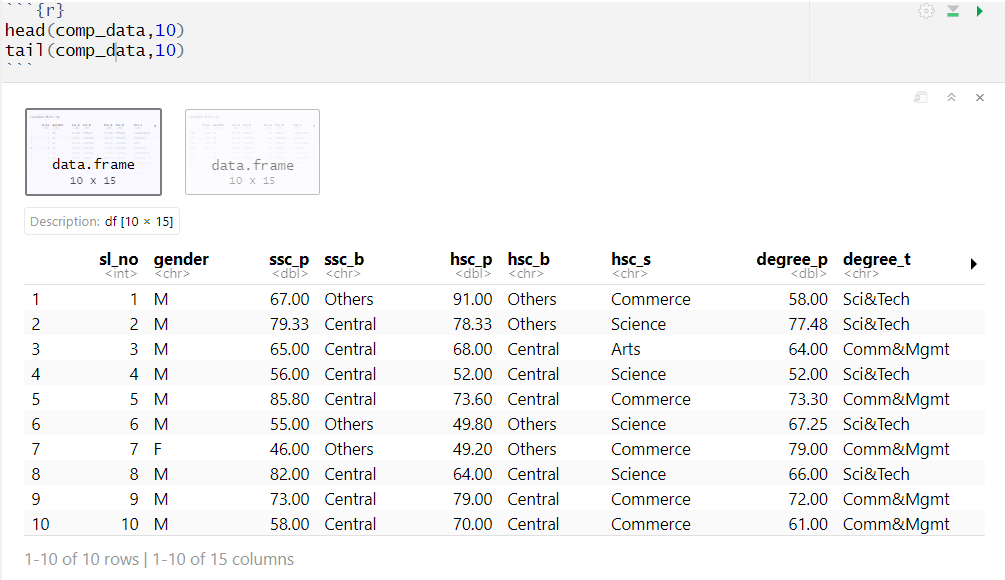


* Select the file that needs to be imported in this case file name is Placement\_Data\_Full\_Class\_2.csv
* The open
* 
* Name the variable in the name section in this case I named it ‘data2’
* Then click Import

**Merge two datasets**

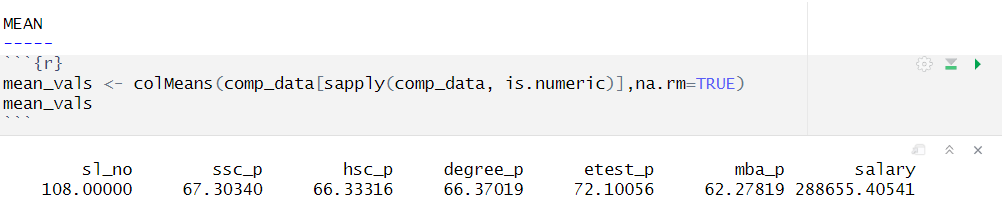
To merge two datasets we use merge function where first parameter is for first dataset second parameter is for second dataset and third parameters is for the column on which merge operation is to be performed.

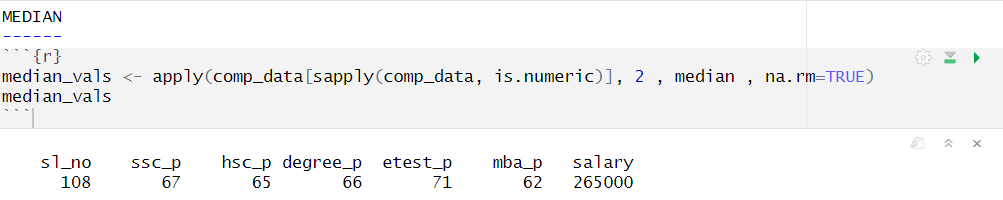
**Show the top 10 and last ten records from dataset with head and tail function**

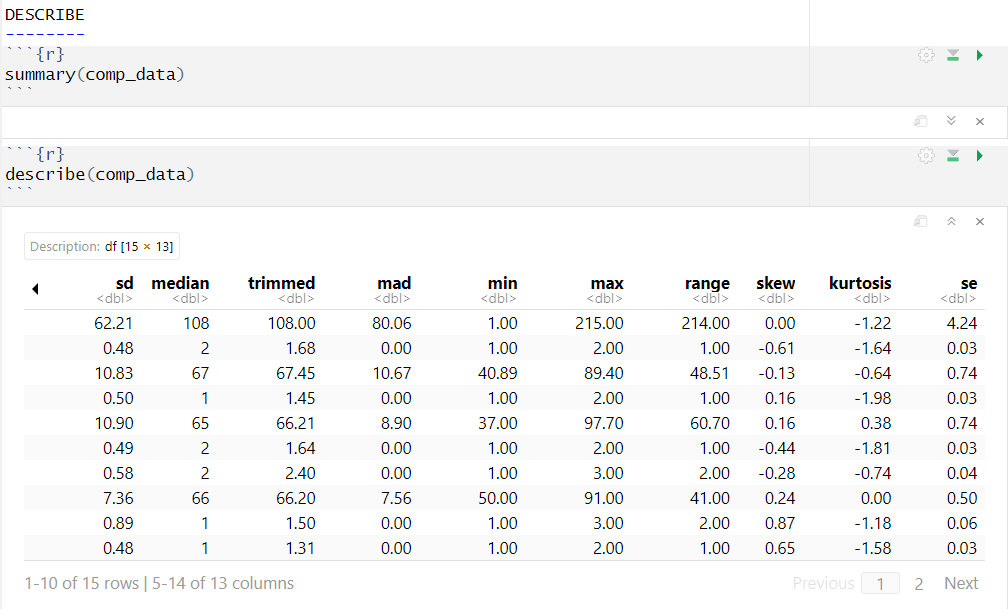
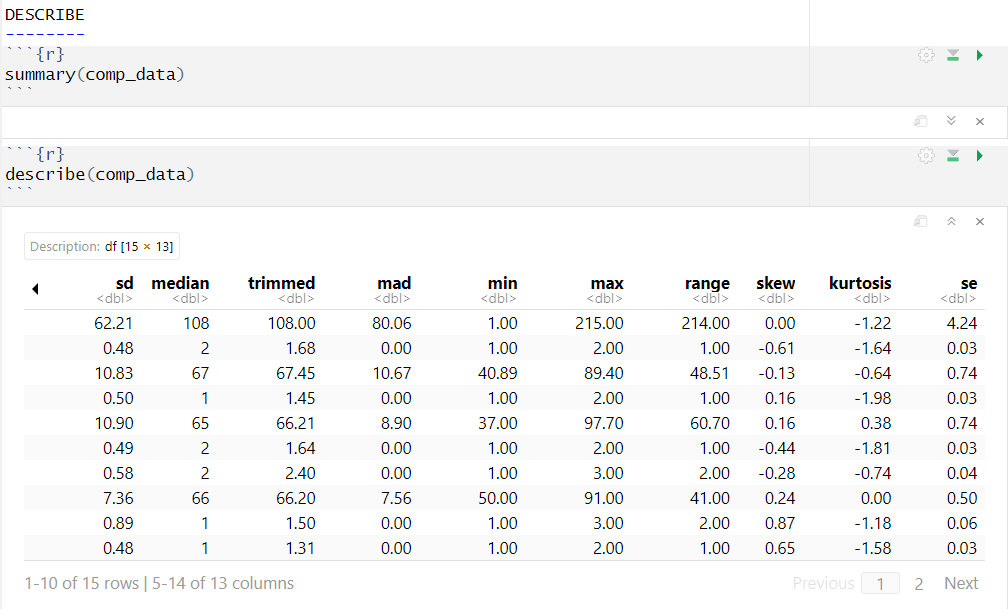
Top 10 records of the data set.

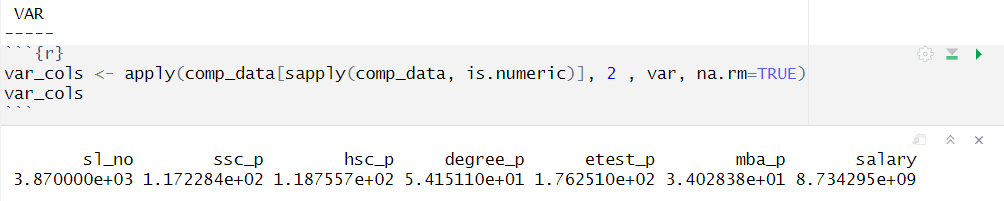
Last 10 records of the dataset.

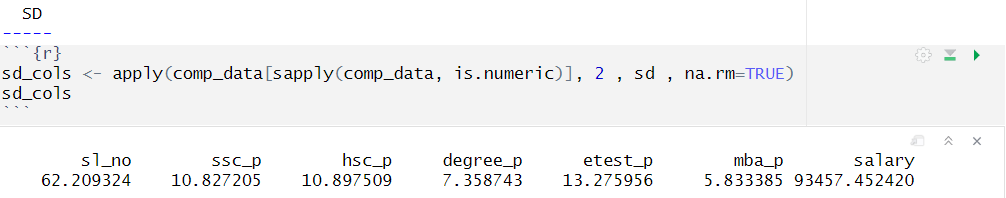
**Implement the following functions**

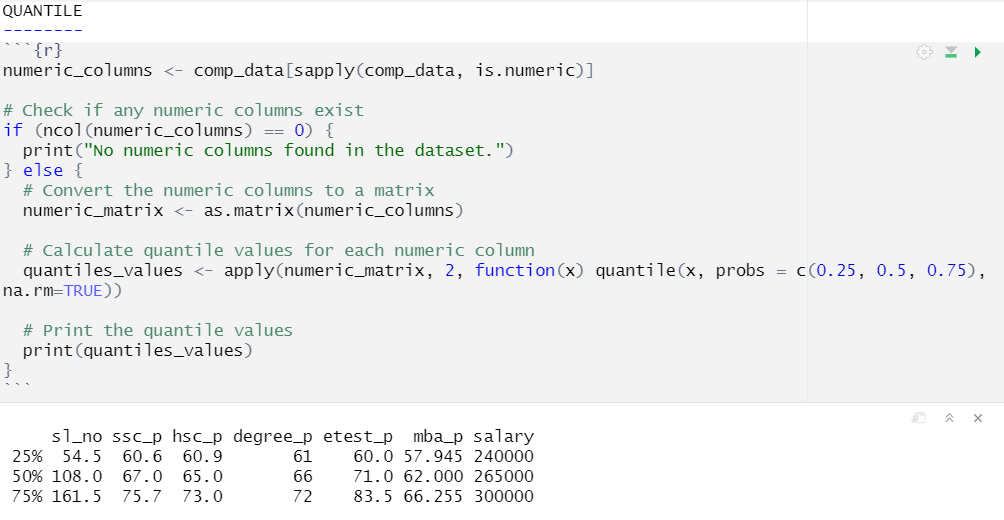
a. MEAN

b. MEDIAN

c. DESCRIBE

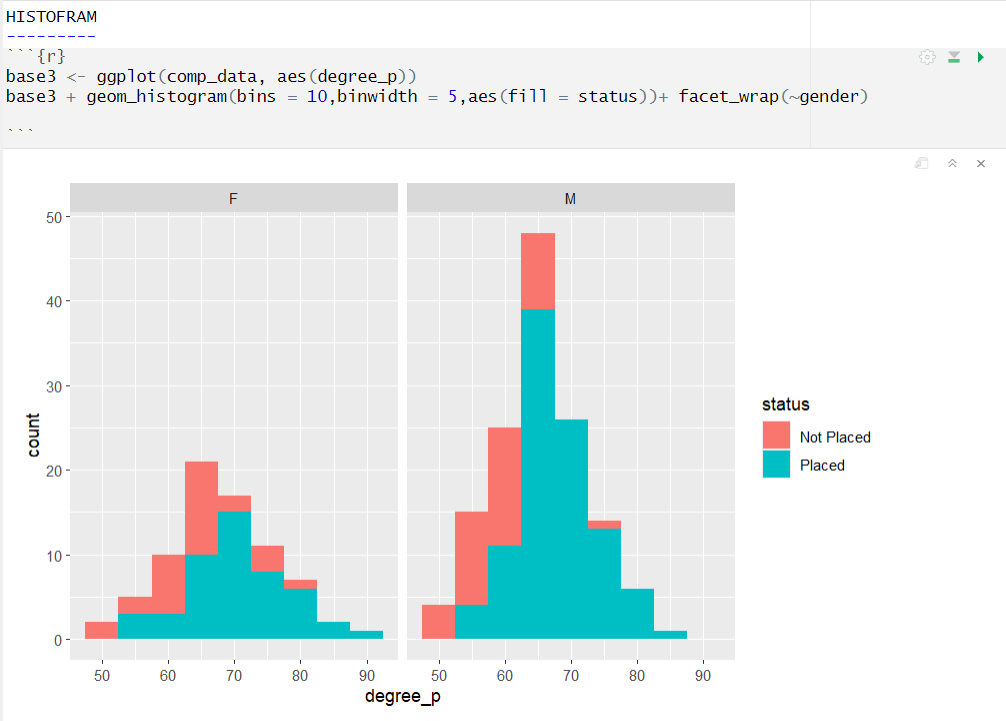
d. VAR(VARIENCE)

e.sd

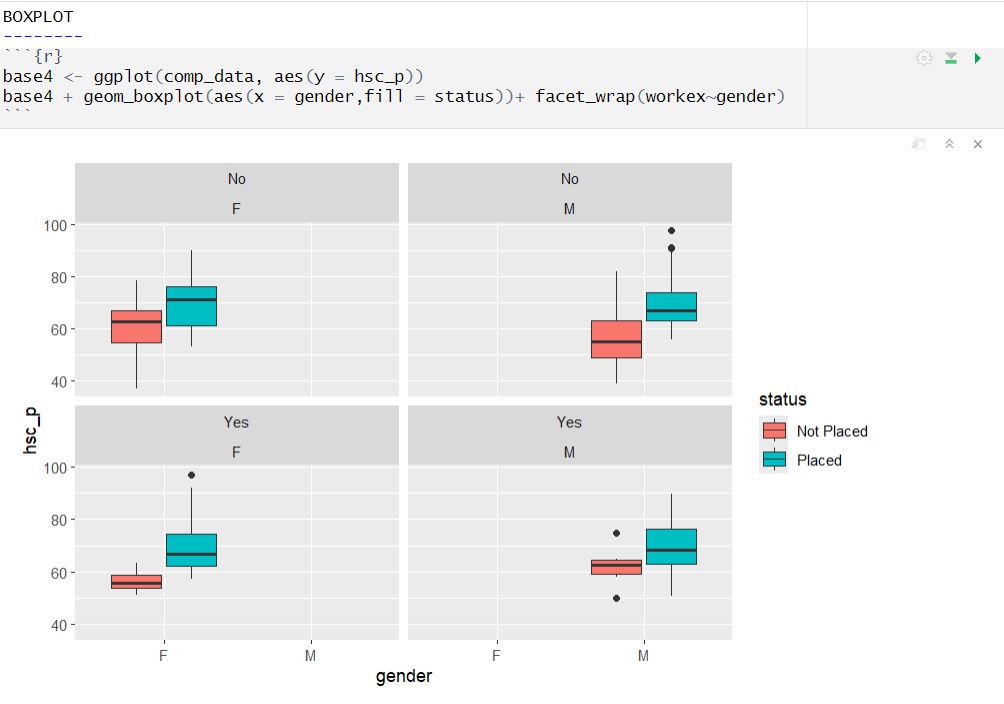
f. Quantile

**Show the followings charts/Graphs on the basis of the dataset**

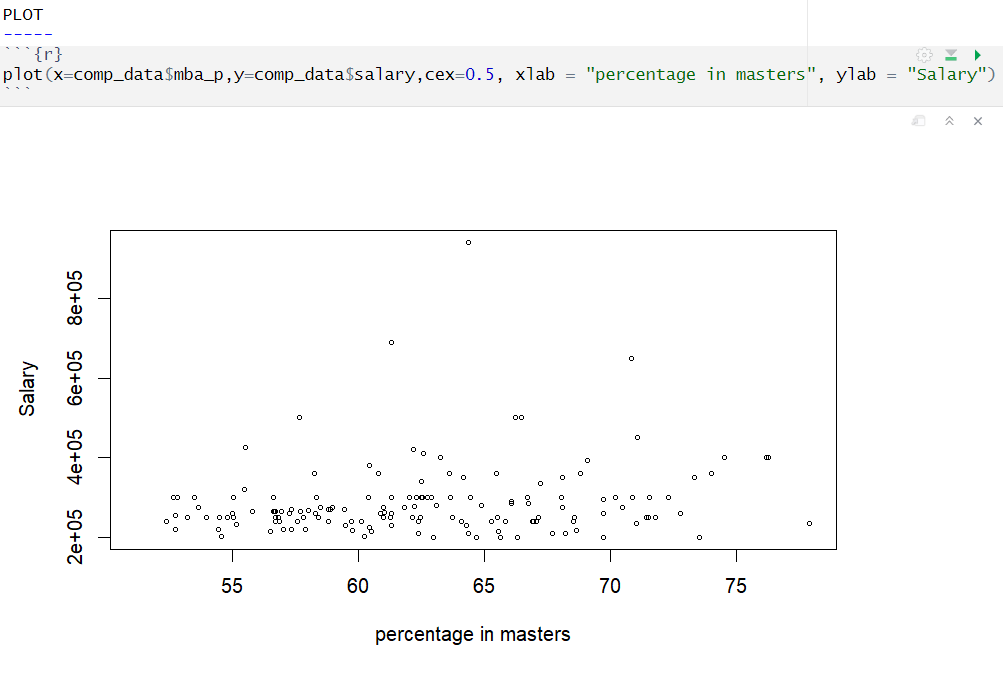
1. HISTOGRAM



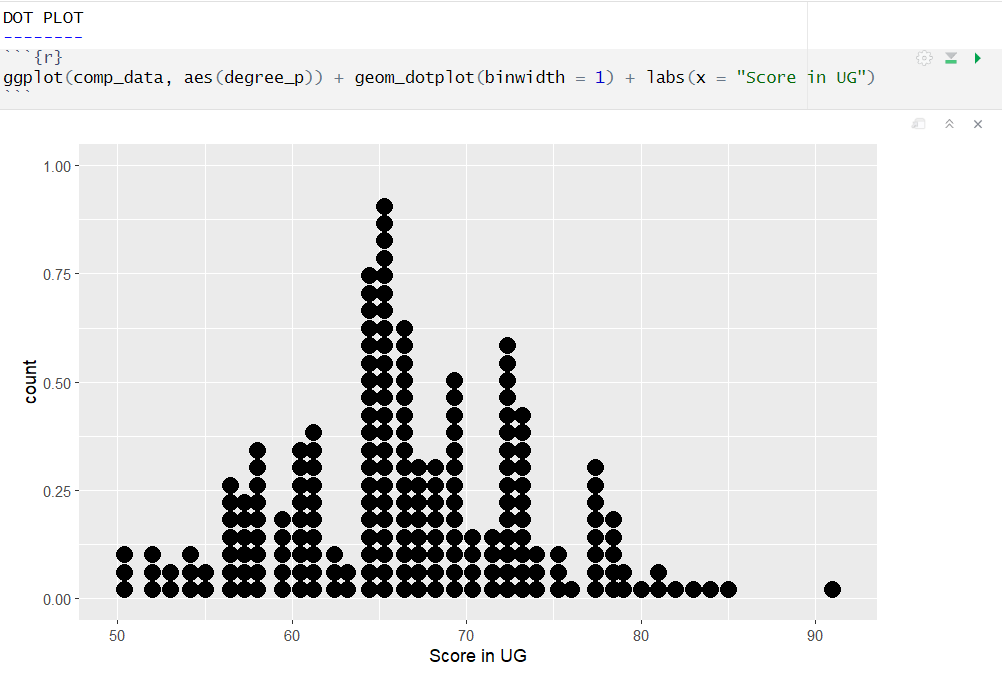
1. BOXPLOT



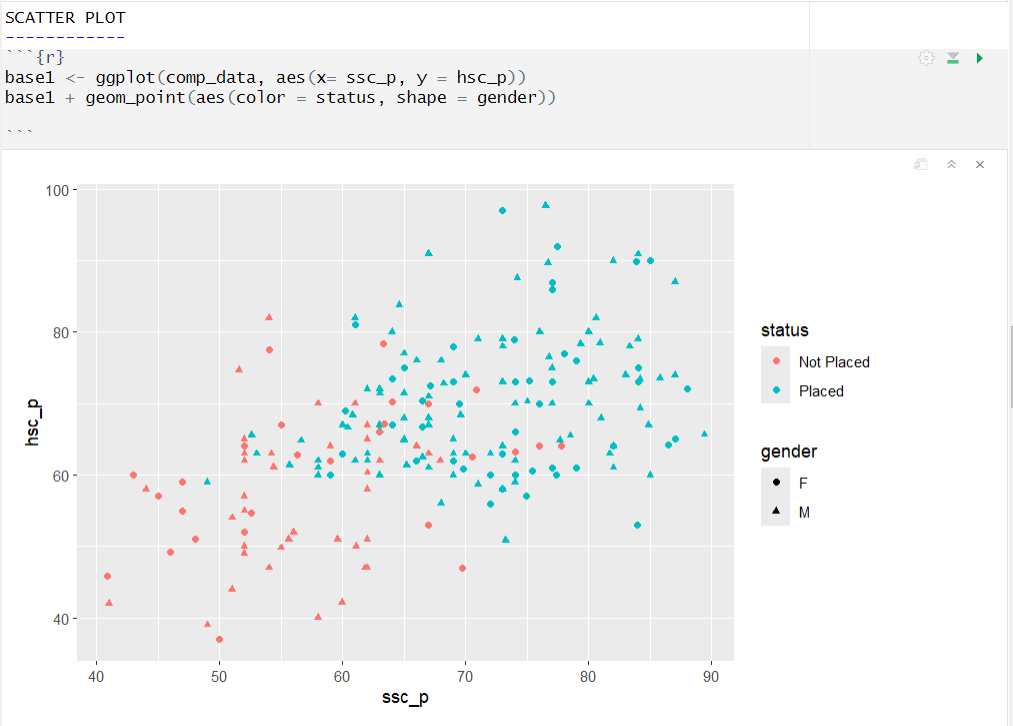
1. PLOT



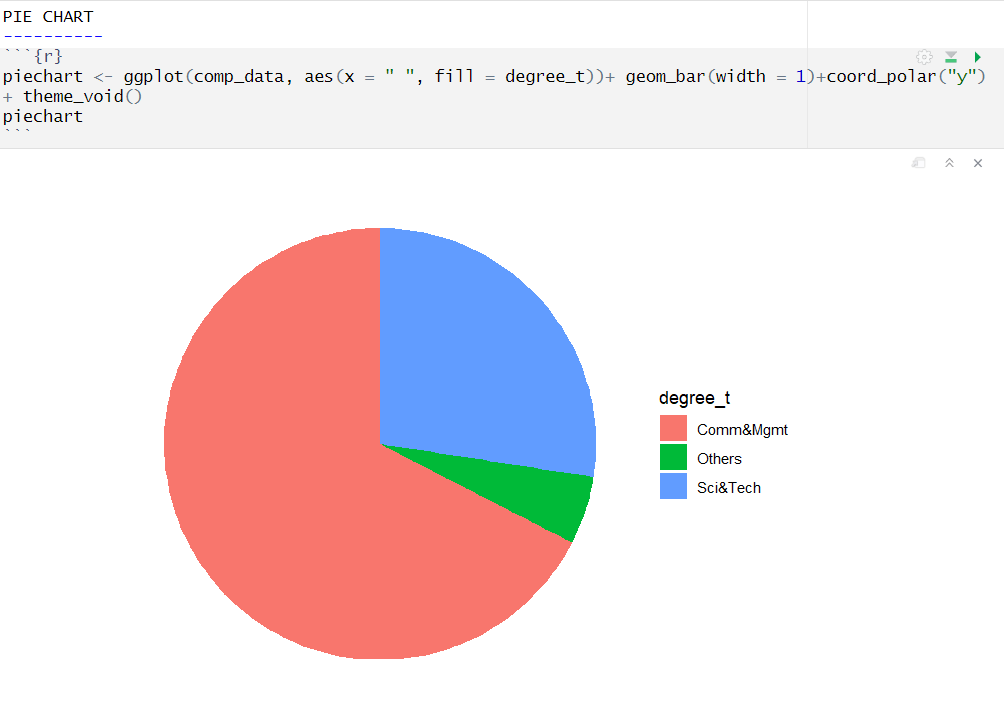
1. DOTPLOT



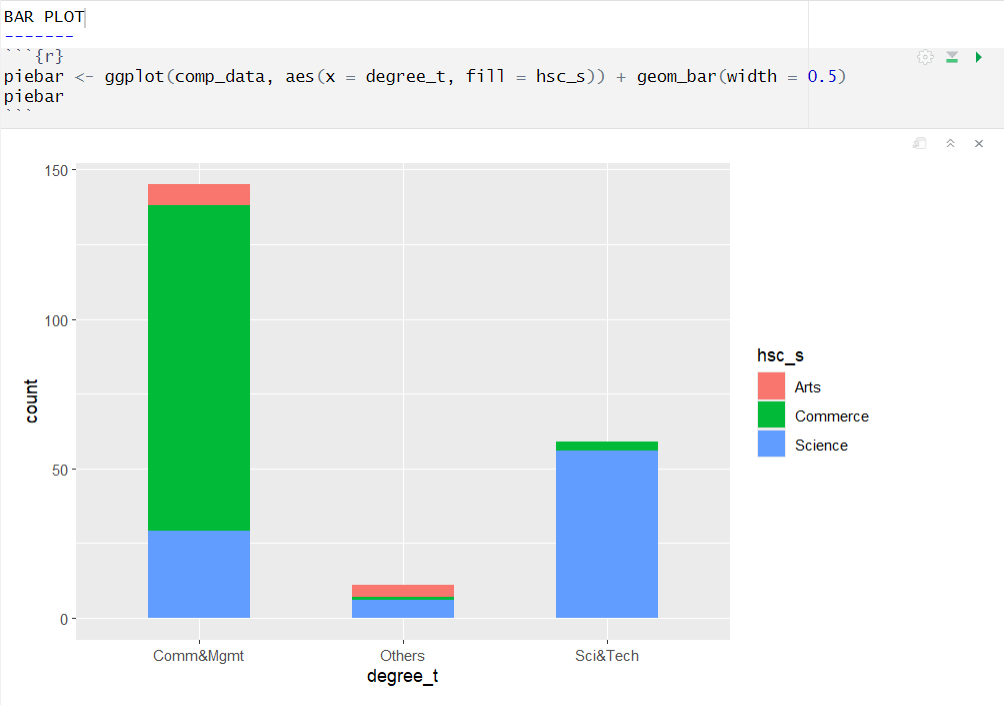
1. SCATTER PLOT



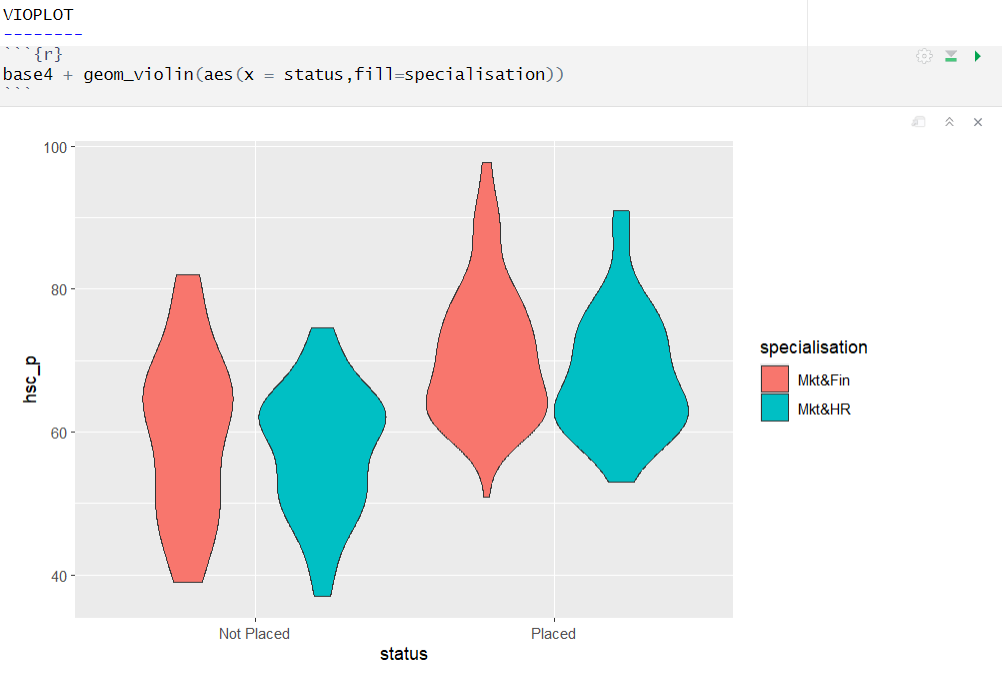
1. PIE PLOT



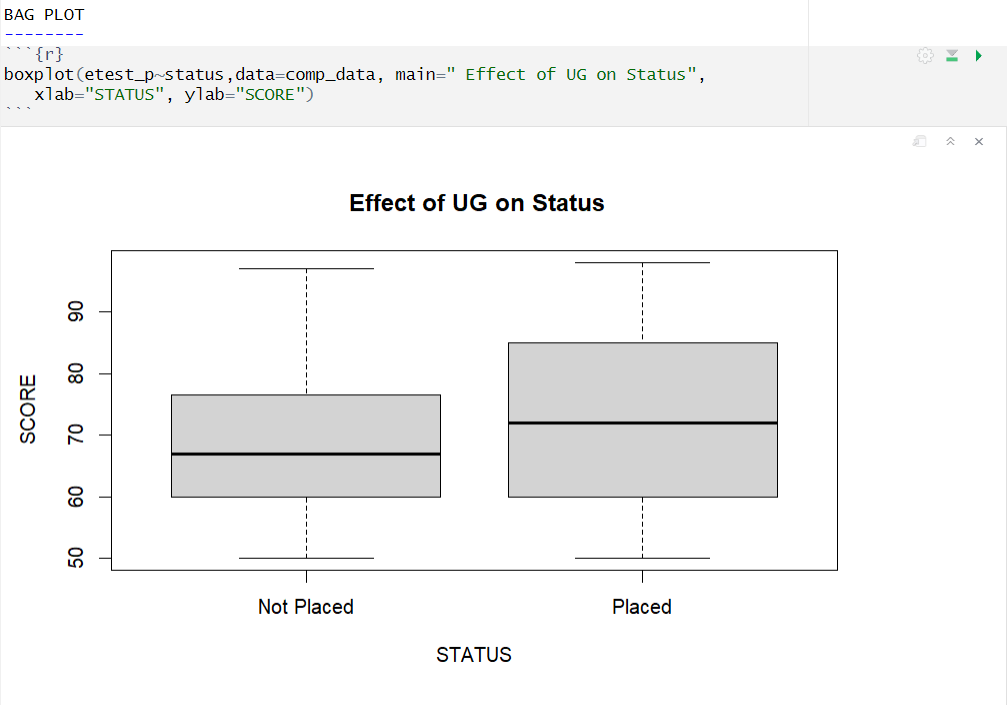
1. BAR PLOT



1. VIOPLOT

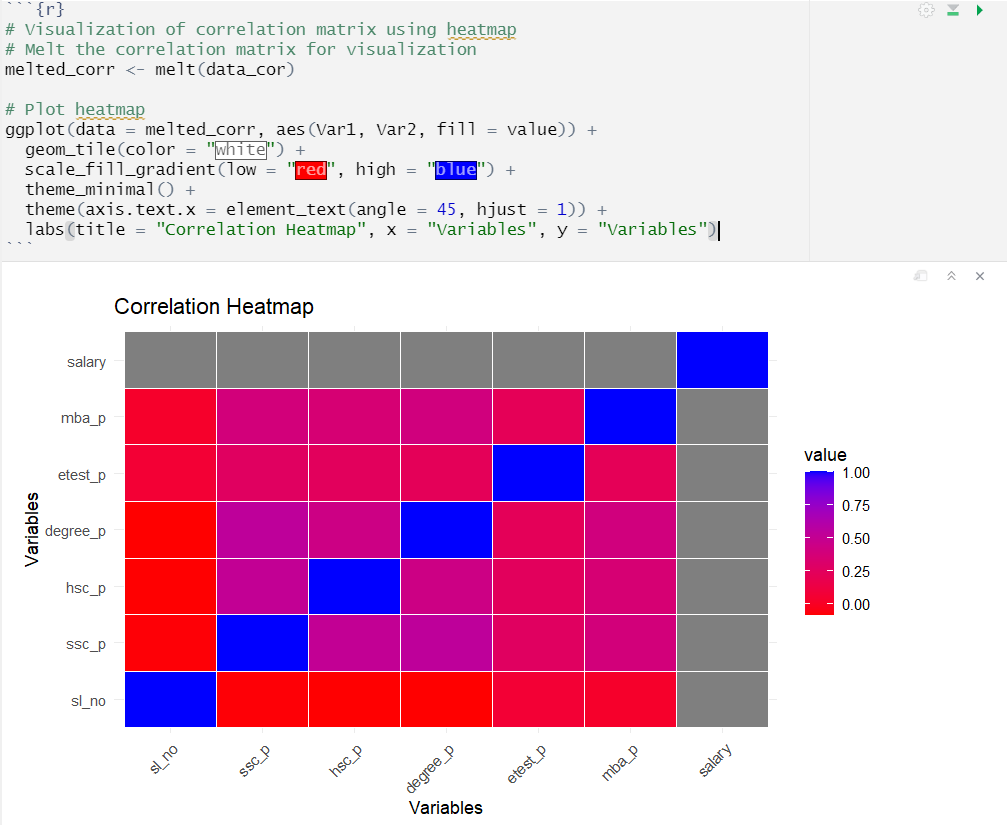
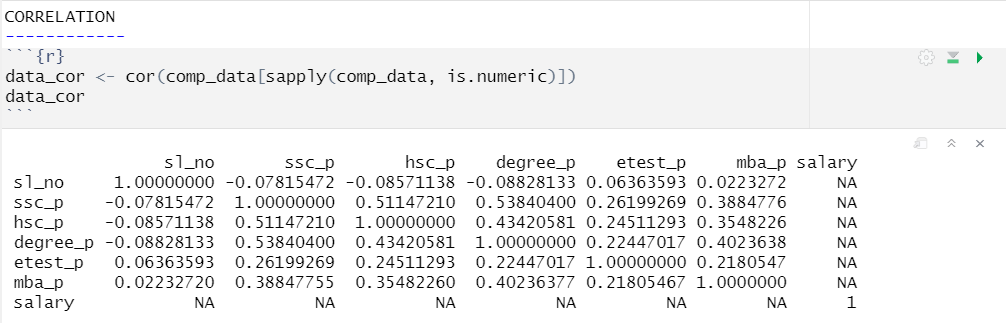


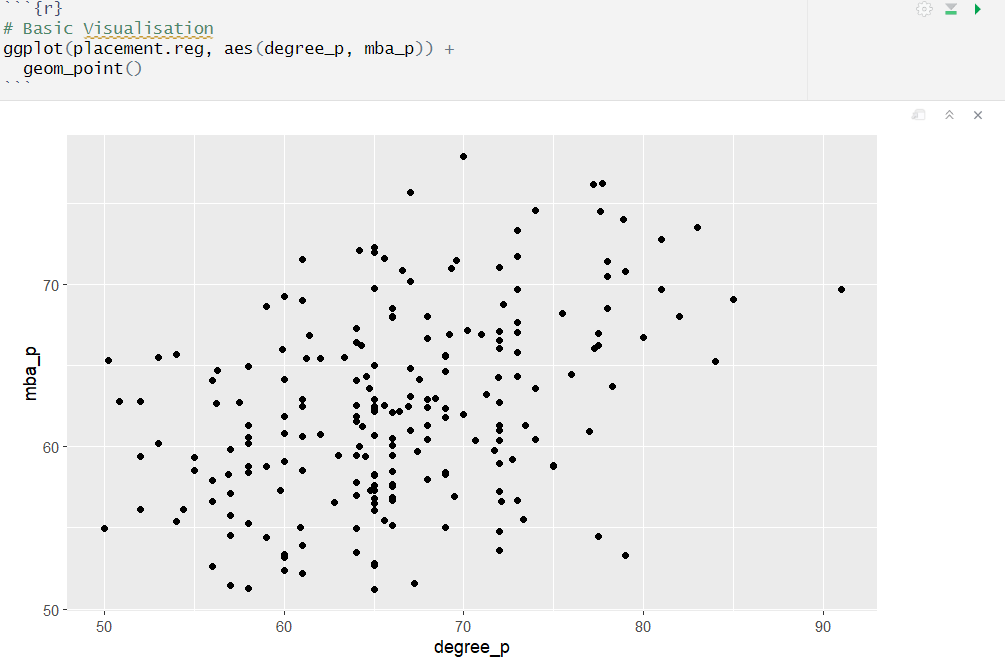
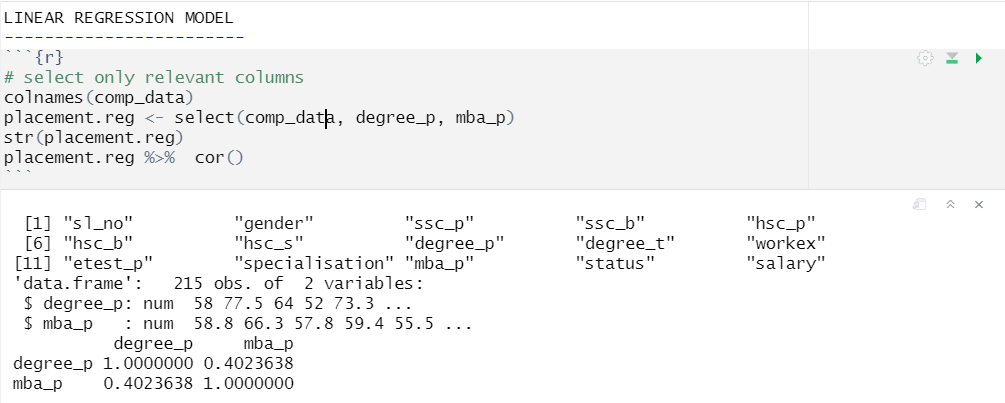
1. BAGPLOT

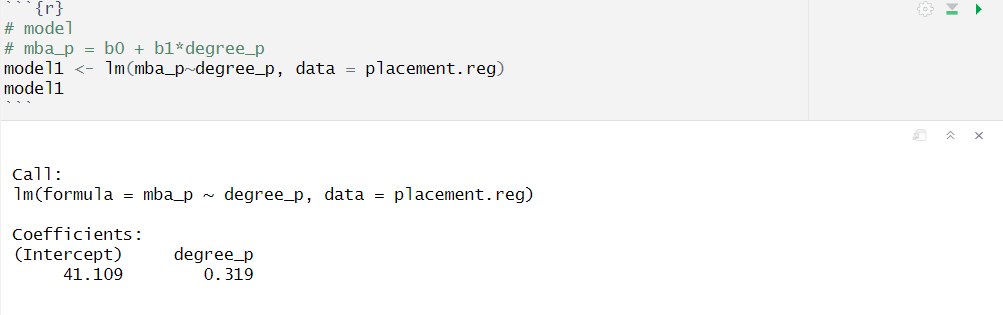
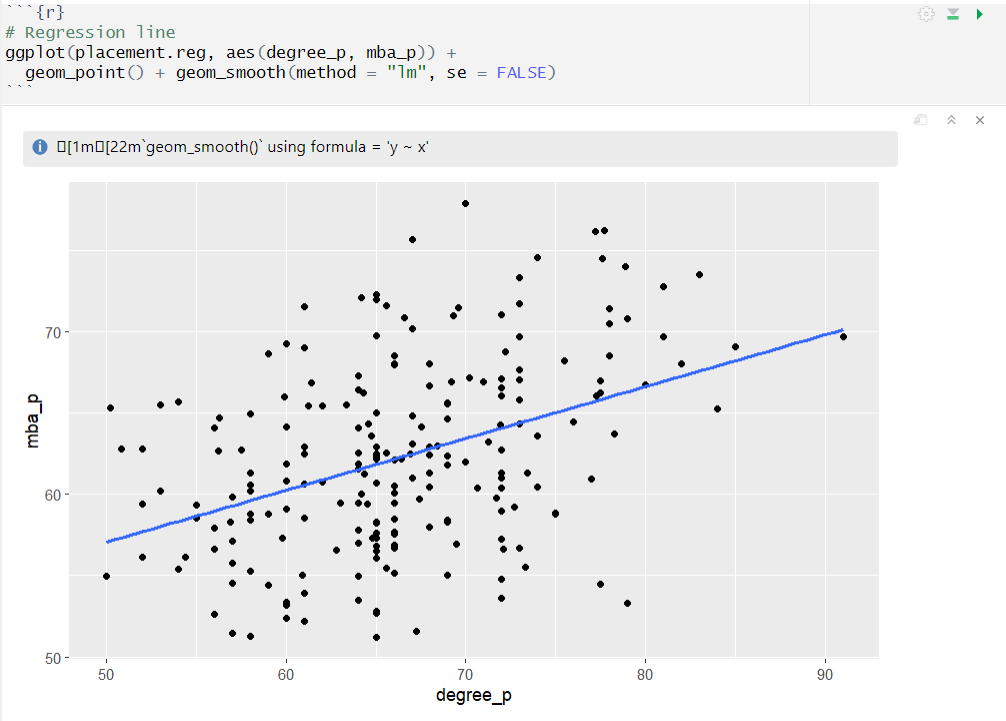
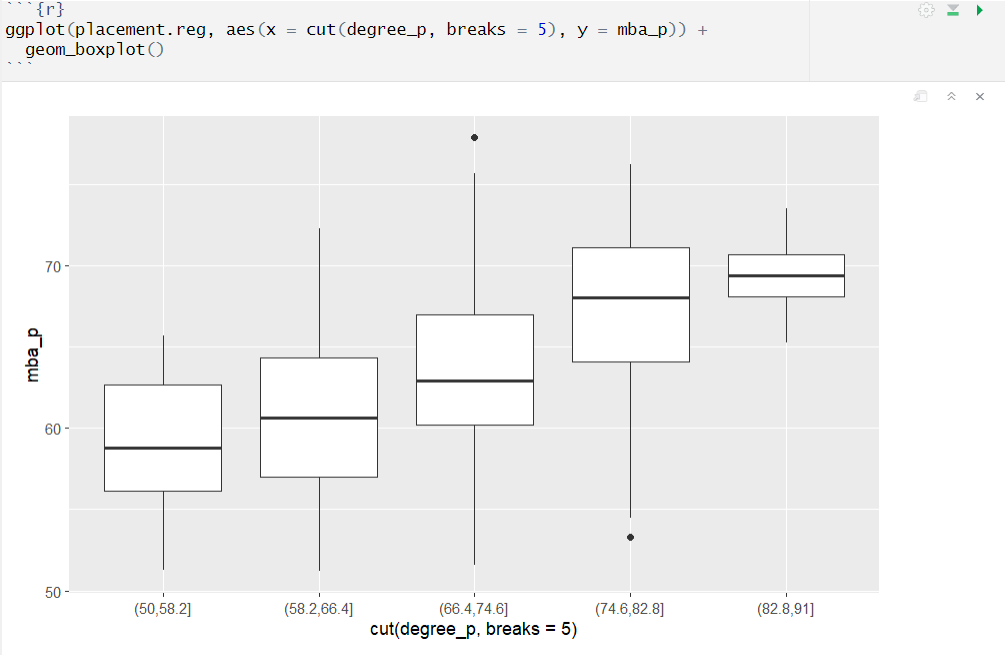
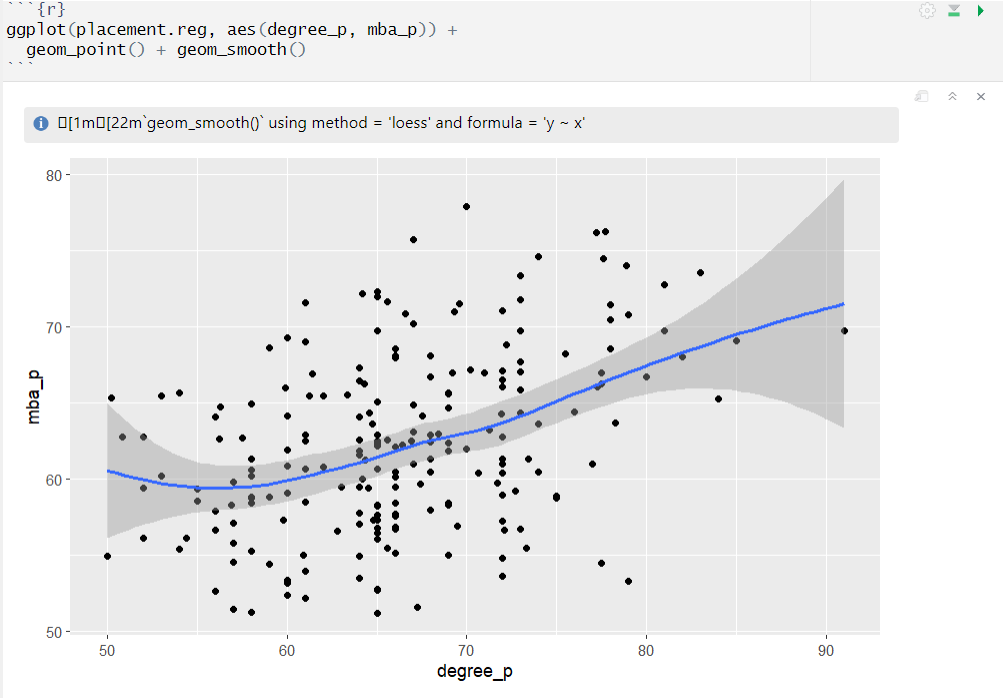


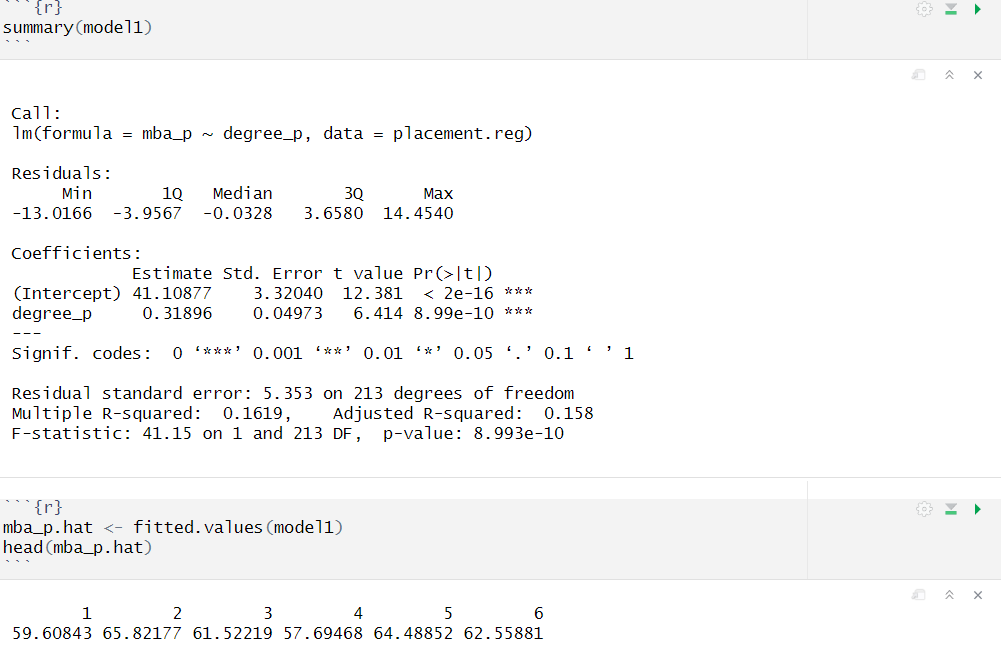
**APPLY FOLLOWING TESTS ON GIVEN DATASET IN R**

**Correlation**

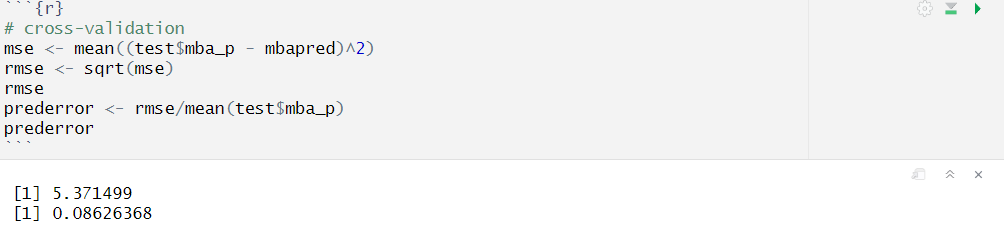
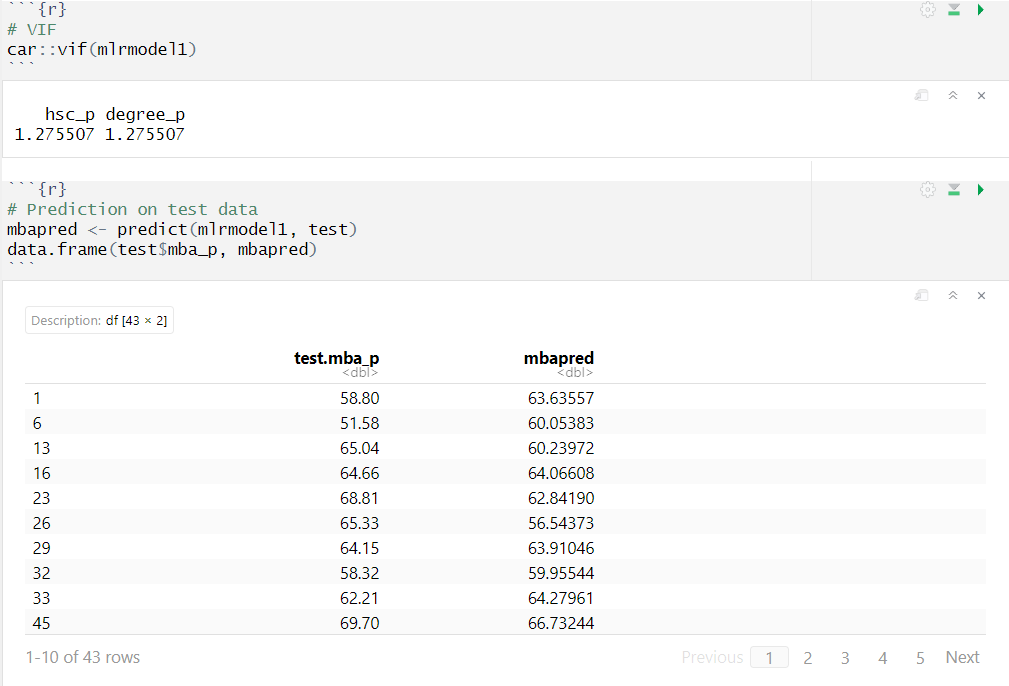
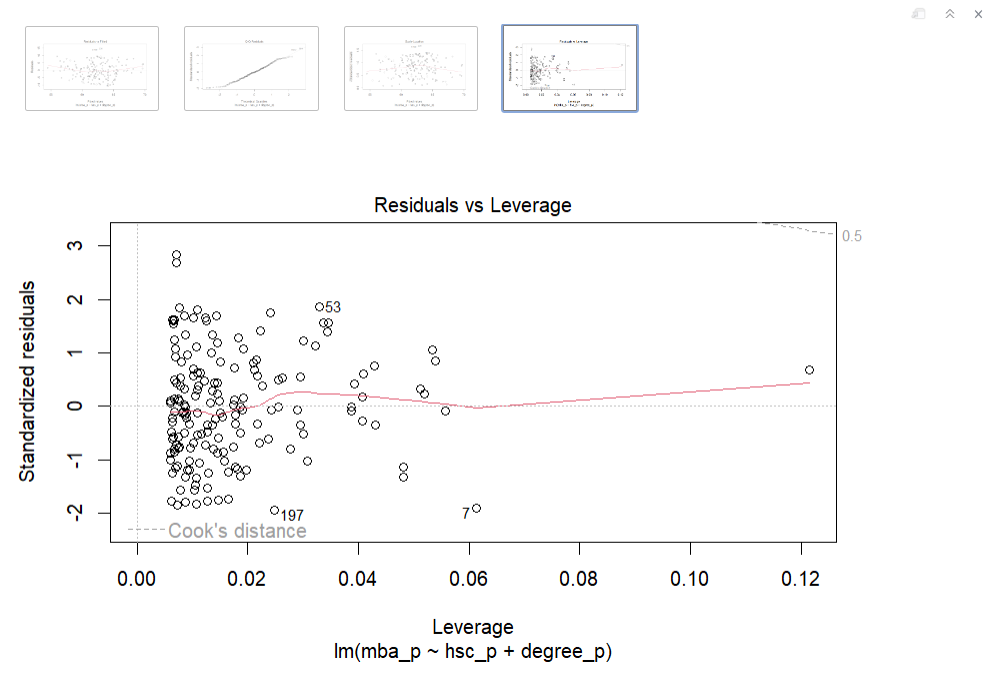
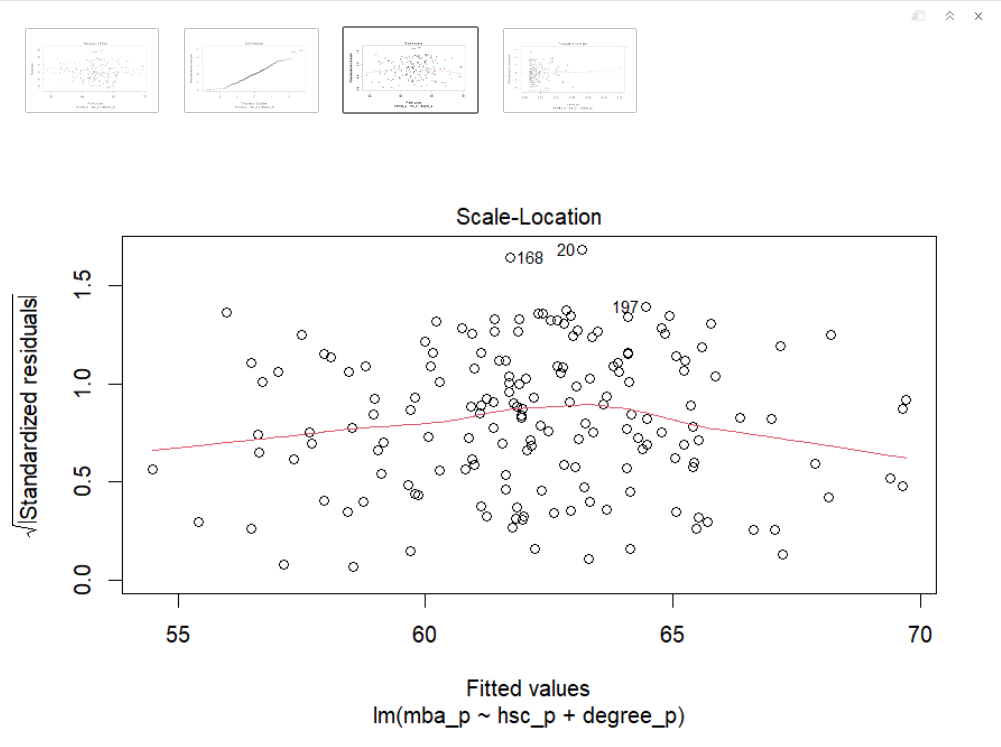
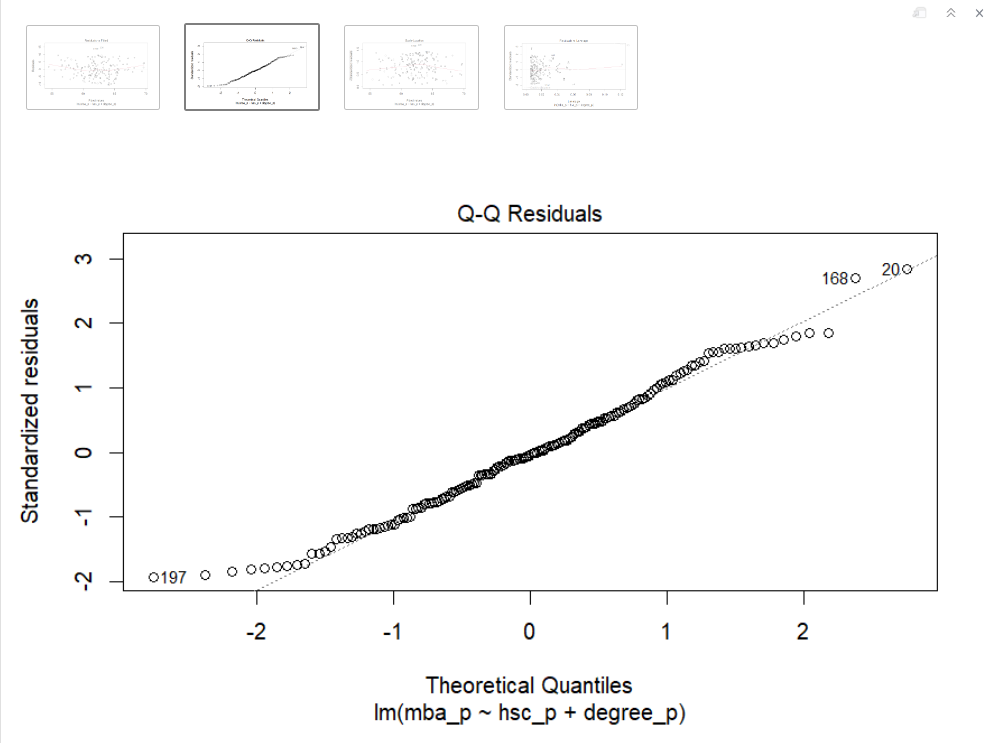
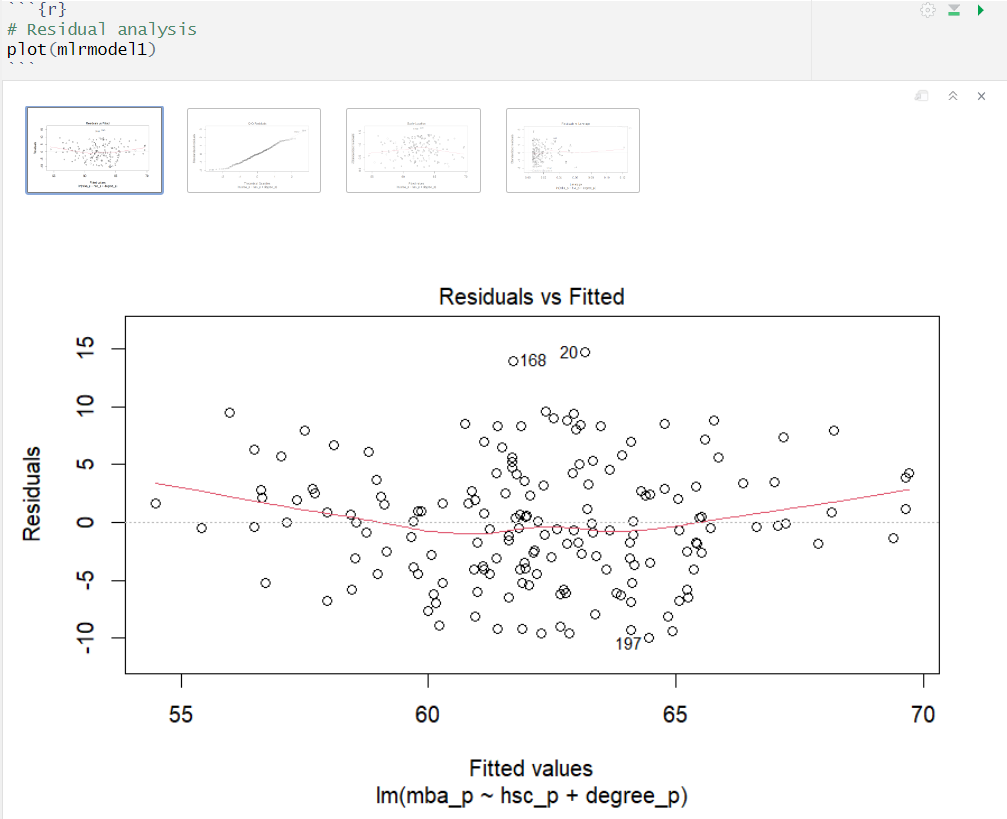
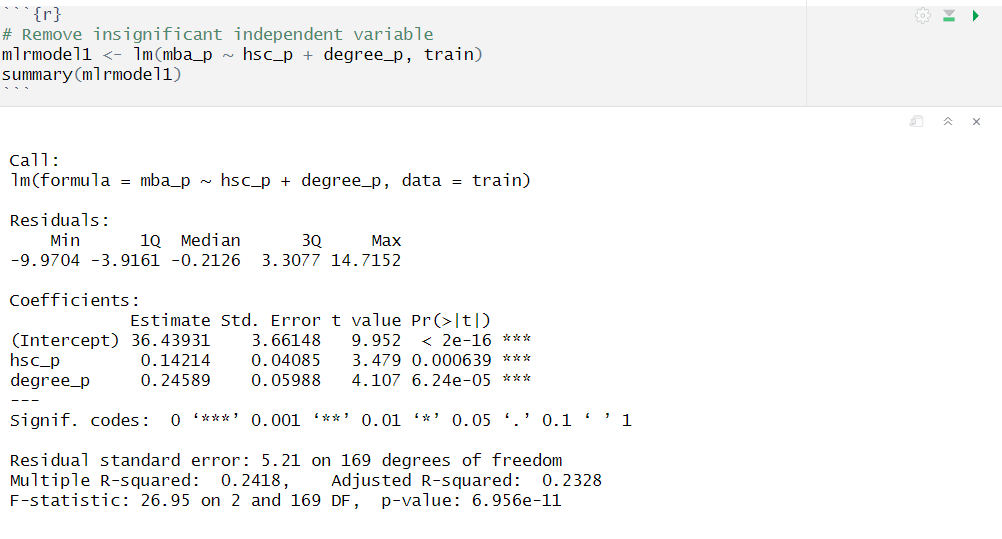
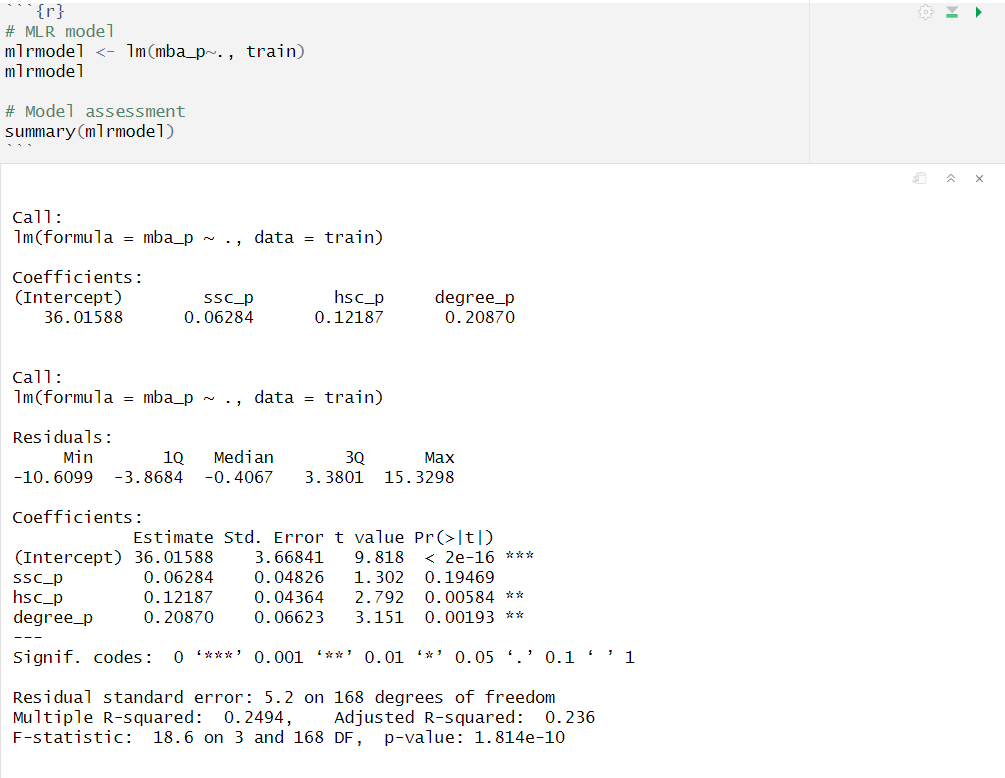
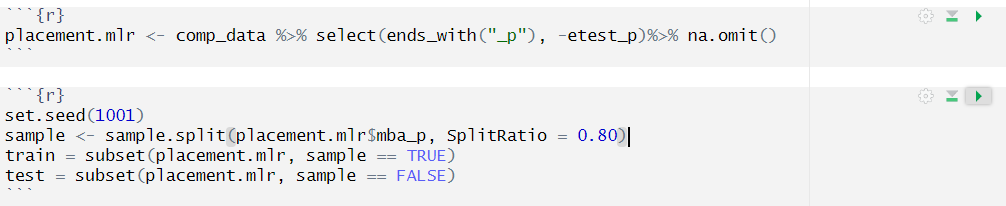
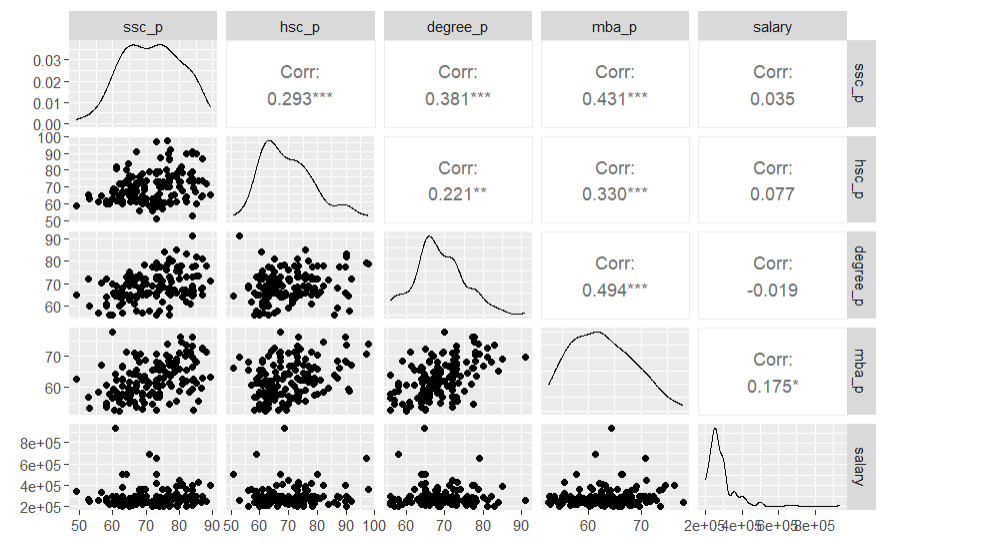
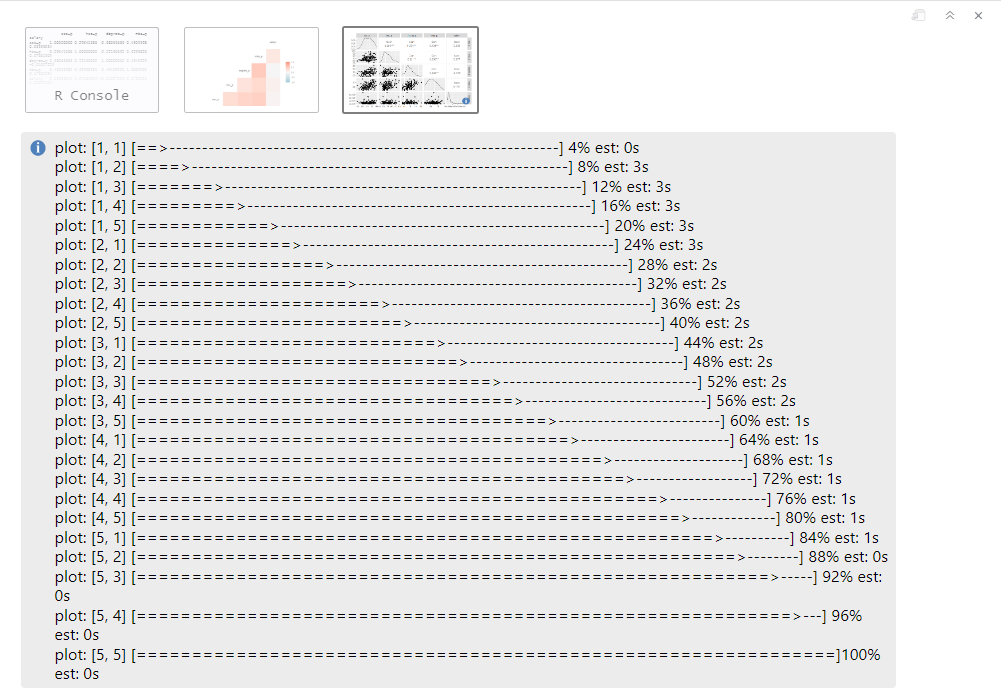
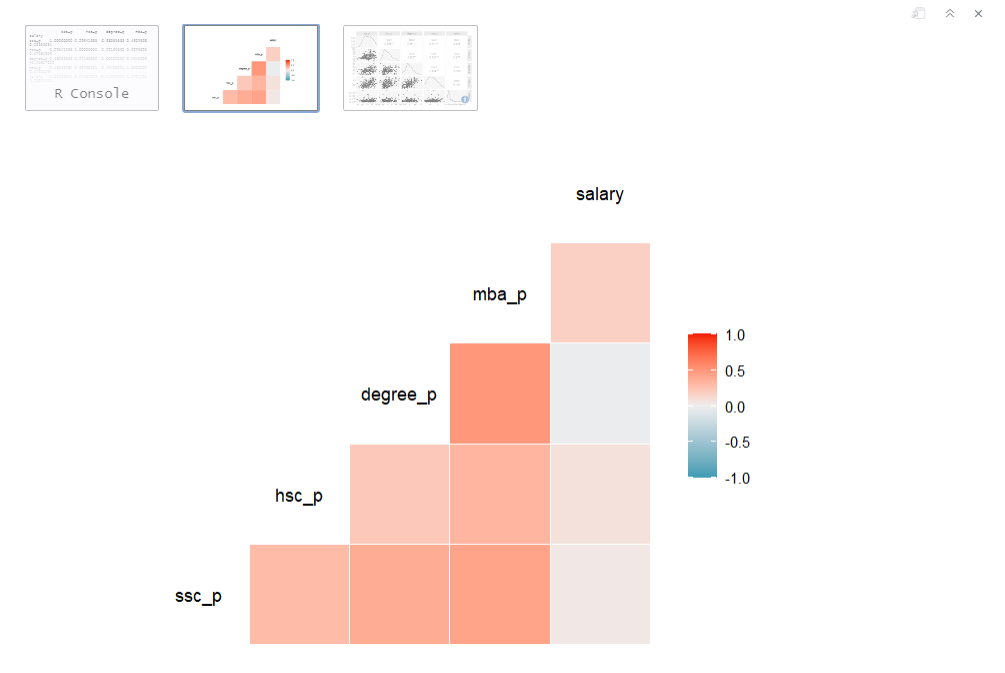
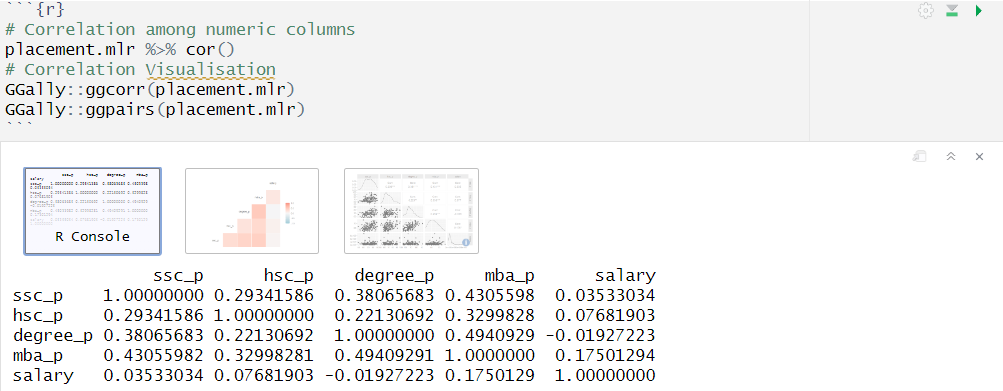
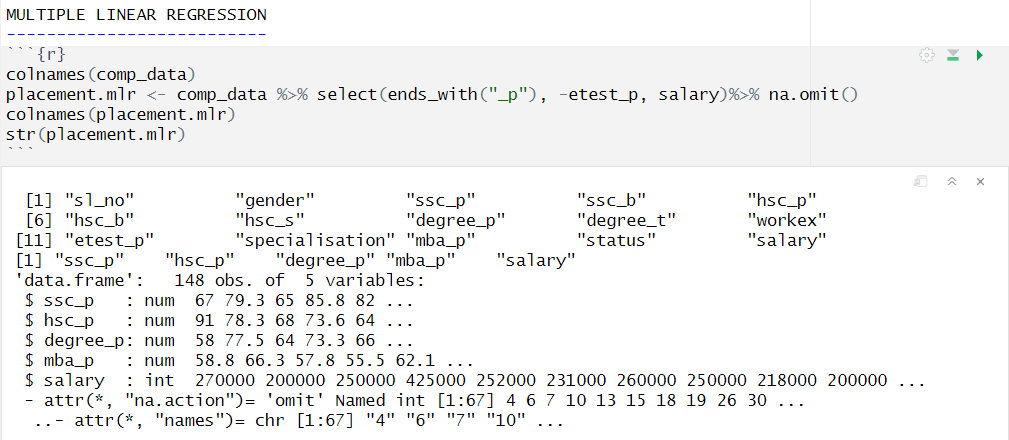


**Linear Regression Model**

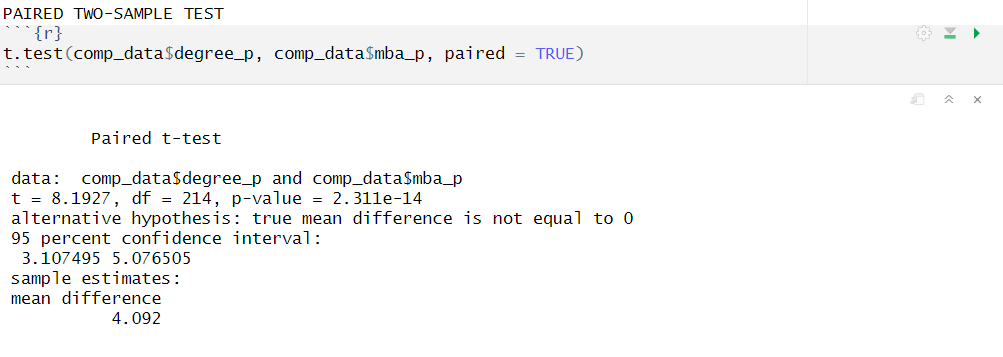
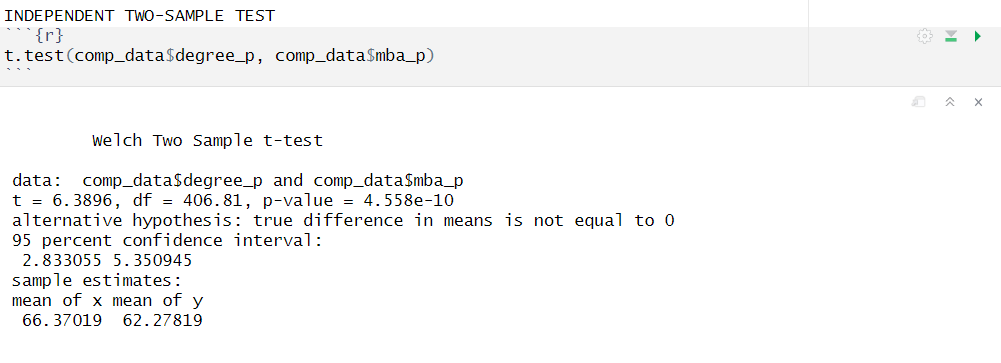
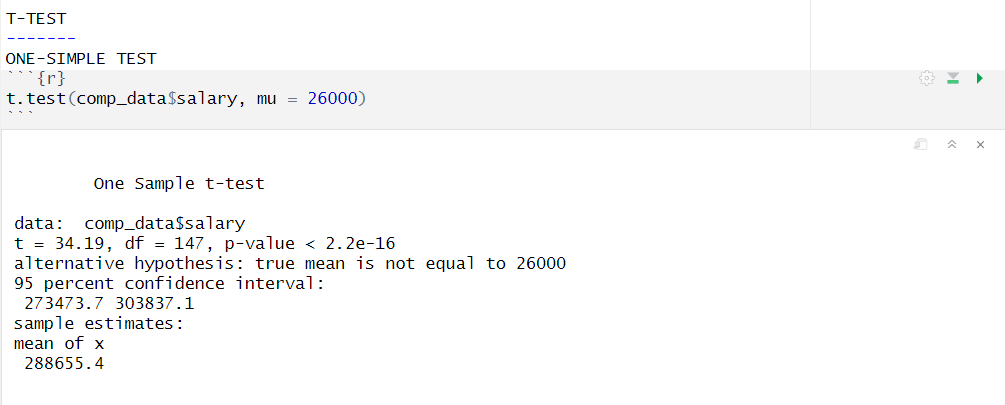




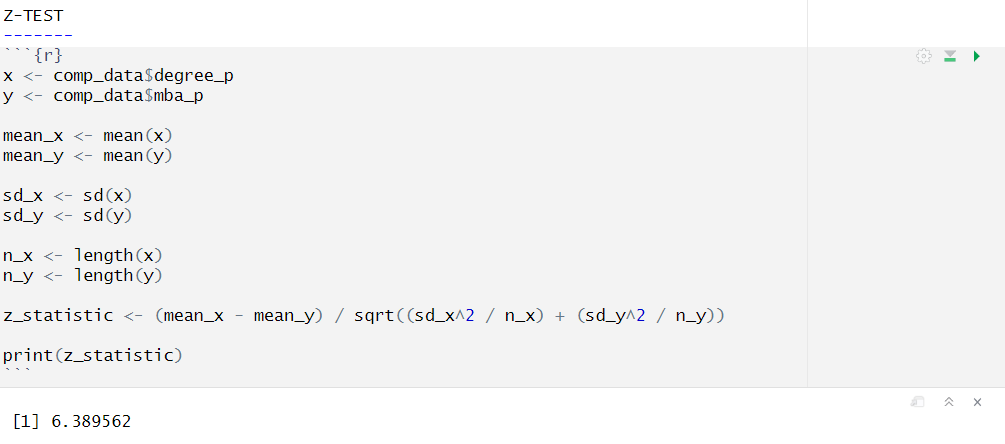
**Multiple Regression Model**

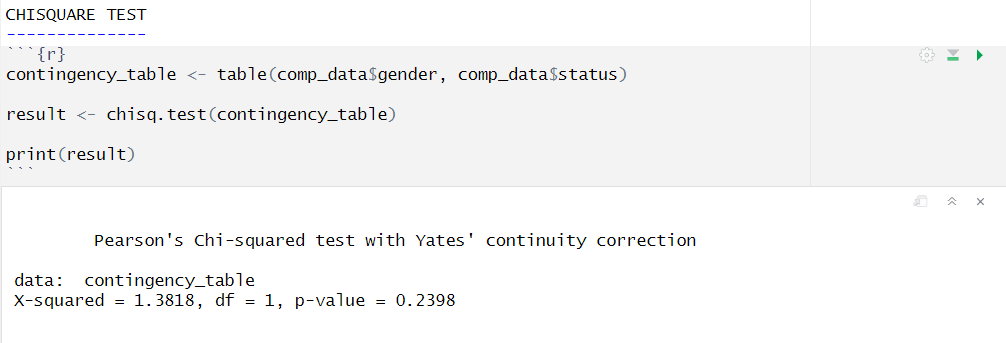


**T-test**



**Z-test**



**Chi-square test**

[**Link to the dataset**](https://www.kaggle.com/datasets/benroshan/factors-affecting-campus-placement)