**ABSTRACT**

Education is the basic requirement for human development. With education, employment opportunities are broadened and income levels are increased. Through our project we aim to understand the differentials and factors associated with school dropouts in India. and it was found that only 75 percent of the children in the age group 6 to 16 years were attending school. About 14 percent of the children never attended the school and 11 percent dropped out of school for various reasons. It was observed that the dropout was high among thechildren belongingregions of Rajasthan and Meghalaya. Parental characteristics also play a significant role in determining school education. The dropout rate was also affected by the various facilities available in the school premises. It was also observed that if parents were not working, the possibility of dropout among their children was relatively high. Through the use of R’s data visualization power we have tried to create a model that will predict the dropout rate and show the reasons of dropout with respect to the dropout rate visually.

* Computer Coding Languages:
  + C++
  + Python
  + Java
  + SQL.
* More Software Engineer Skills:
  + Linux
  + Operating System
  + Data Structures
  + DBMS

**CHAPTER: 1**

**INTRODUCTION**

R is a programming language and software environment for statistical analysis, graphics representation and reporting. R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team.

The core of R is an interpreted computer language which allows branching and looping as well as modular programming using functions. R allows integration with the procedures written in the C, C++, .Net, Python or FORTRAN languages for efficiency.

R is freely available under the GNU General Public License, and pre-compiled binary versions are provided for various operating systems like Linux, Windows and Mac. R is free software distributed under a GNU-style copy left, and an official part of the GNU project called **GNU S**.

* 1. **Evolution of R**

R was initially written by **Ross Ihaka** and **Robert Gentleman** at the Department of Statistics of the University of Auckland in Auckland, New Zealand. R made its first appearance in 1993.

• A large group of individuals has contributed to R by sending code and bug reports.

• Since mid-1997 there has been a core group (the "R Core Team") who can modify the R source code archive.

**1.2 Features of R**

As stated earlier, R is a programming language and software environment for statistical analysis, graphics representation and reporting. The following are the important features of R:

• R is a well-developed, simple and effective programming language which includes conditionals, loops, user defined recursive functions and input and output facilities.

• R has an effective data handling and storage facility,

• R provides a suite of operators for calculations on arrays, lists, vectors and matrices.

• R provides a large, coherent and integrated collection of tools for data analysis.

• R provides graphical facilities for data analysis and display either directly at the computer or printing at the papers.

R is a powerful data analysis and statistical tool which helps in understanding the data in a much more consolidated and better manner. It provides us with various inbuilt functions that help in the visualization of data and also help in drawing hidden facts and conclusions from the data.

The various libraries, packages and functions that we have used in our project are:

1. **Package: ggplot2 / ggthemes**

The ggplot2 package, created by Hadley Wickham, offers a powerful graphics language for creating elegant and complex plots., ggplot2 allows you to create graphs that represent both univariate and multivariate numerical and categorical data in a straightforward manner. Grouping can be represented by color, symbol, size, and transparency. A ggplot is made up of two main components – a ggplot() object and at least one geom layer. The various layers of a ggplot are:

* Data
* Aesthetics
* Geometrics
* Facets
* Statistics
* Scales

Other than this the various functions that have been used with geom are geom\_point(),geom\_smooth().

1. **Package: corrgram / corrplot**

Correlograms help us visualize the data in correlation matrices. In R, correlograms are implemented through the corrgram(x, order = , panel=, lower.panel=, upper.panel=, text.panel=, diag.panel=) function in the corrgram package.

* x is a data frame with one observation per row.
* order=TRUE will cause the variables to be ordered using principal component analysis of the correlation matrix.
* panel= refers to the off-diagonal panels. You can use lower.panel= and upper.panel= to choose different options below and above the main diagonal respectively. text.panel= and diag.panel= refer to the main diagnonal.

1. **Package: dplyr**

dplyr is a powerful R-package to transform and summarize tabular data with rows and columns. The package contains a set of functions (or “verbs”) that perform common data manipulation operations such as filtering for rows, selecting specific columns, re-ordering rows, adding new columns and summarizing data.In addition, dplyr contains a useful function to perform another common task which is the “split-apply-combine” concept. Compared to base functions in R, the functions in dplyr are easier to work with, are more consistent in the syntax and are targeted for data analysis around data frames instead of just vectors.

1. **Package: caTools**

The functions in this particular package are used mainly for training the data, generating models and further making predictions with those models. Then the error can also calculated between the predicted values and the original values. The various functions used under caTools are- split function, train function etc.

**CHAPTER :2**

**DROPOUT RATE ANALYSIS**

Student dropout is a major concern in the education and policy-making communities. About 32% of students seeking bachelor’s degrees do not complete their degree within 6 years and 27% of full-time first-year students seeking baccalaureate degrees who do not return for a second year. Despite long-standing theory, student drop out continues to be a large concern to the education community and policy makers. Our broader objective is to understand the key determinants of dropout. Our work relates to analyzing the reasons of dropouts, mainly from primary schools.

•To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India.

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

To understand the differentials in school dropouts at national and at state levels.

•To understand the main reasons for school dropout reported by household members.

•To examine the household and parental characteristics influencing the school dropouts in India

* 1. **Need for the study**

Despite of governmental efforts, huge investment and many innovative programs, the school dropout remains alarming high in many states. In this context, it is not only the provision of schooling facilities and quality of education, but also other household and social factors play a major role in influencing the discontinuation of education. It is very important to understand the family and parental characteristics to examine the reasons behind school dropouts. Though micro level studies and research based on departmental statistics have been undertaken by many scholars, National Family Health Survey -3 provides an opportunity to examine the school dropout, based on large nationally representative sample survey of households in Indian context.

* 1. **Objectives**
* To understand the differentials in school dropouts at national and at state levels.
* To understand the main reasons for school dropout reported by household members.
* To examine the household and parental characteristics influencing the school dropouts in India

**2.3Data Source:**

Data for this study has been obtained from data.gov.in. The available data on school dropout of children were analyzed to find out the extent of school dropout, stagesof school dropout, household and parentalcharacteristics ofchildren who dropped out, and the reported reasons for discontinuing the school education.

**CHAPTER 3**

**ABOUT THE PROJECT**

The data obtained from data.gov.in was put in an excel file and then converted to a comma separated file.

The data is as follows:

Column Names:

1. Dropout.Rate
2. With.Pucca.building
3. With.Boundary.wall
4. With.Drinking.water
5. With.Common.toilet
6. With.toilets.for.girls
7. With.Blackboard
8. With.Electricity
9. With.Computer.centre
10. With.TLM
11. With.1.3.classrooms

Dimensions: 11 rows and 11 columns

The data obtained from data.gov.in was put in an excel file and saved in a .csv format like so:

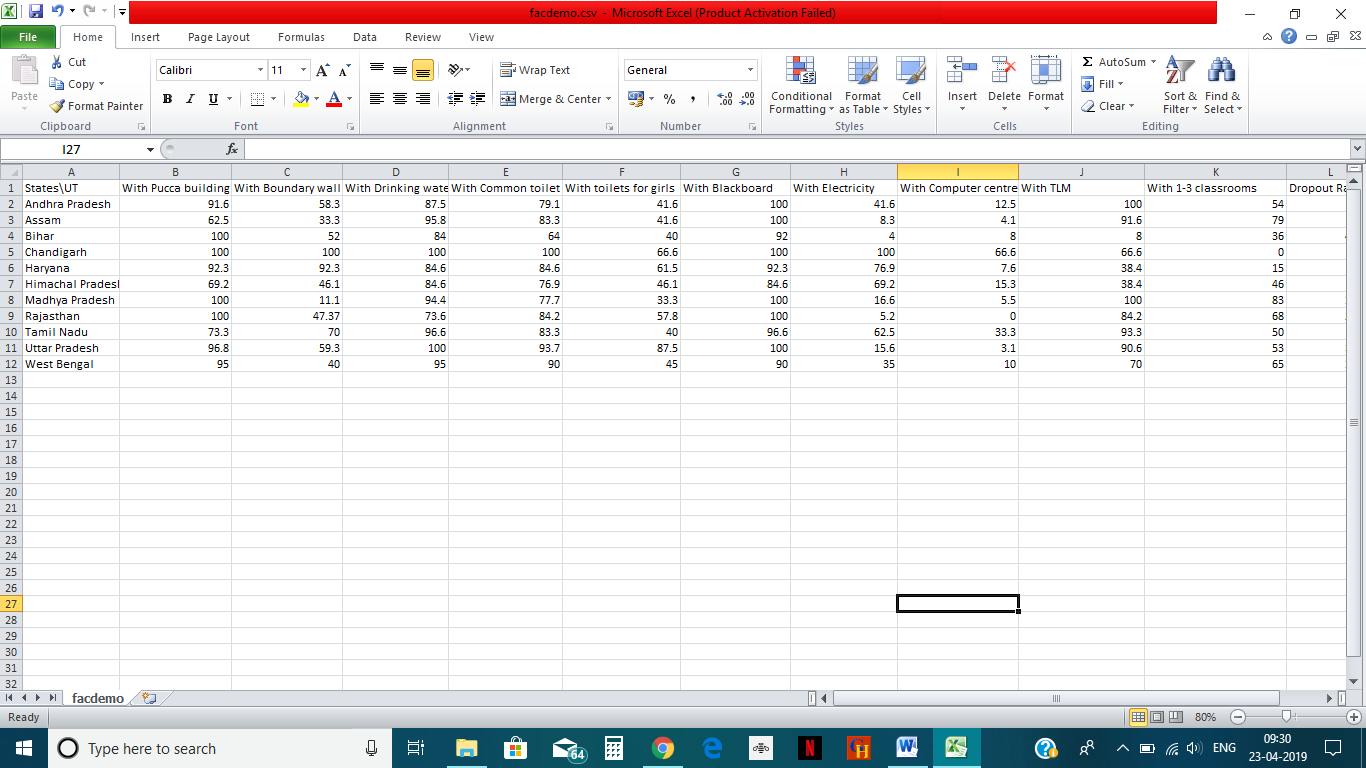


Fig 1: data obtained

After this, the data that was extracted was put in a data frame namely, df5. After assigning the data frame and installing the above mentioned packages and libraries the data analysis was performed.

First a correlation matrix between various matrices was made:

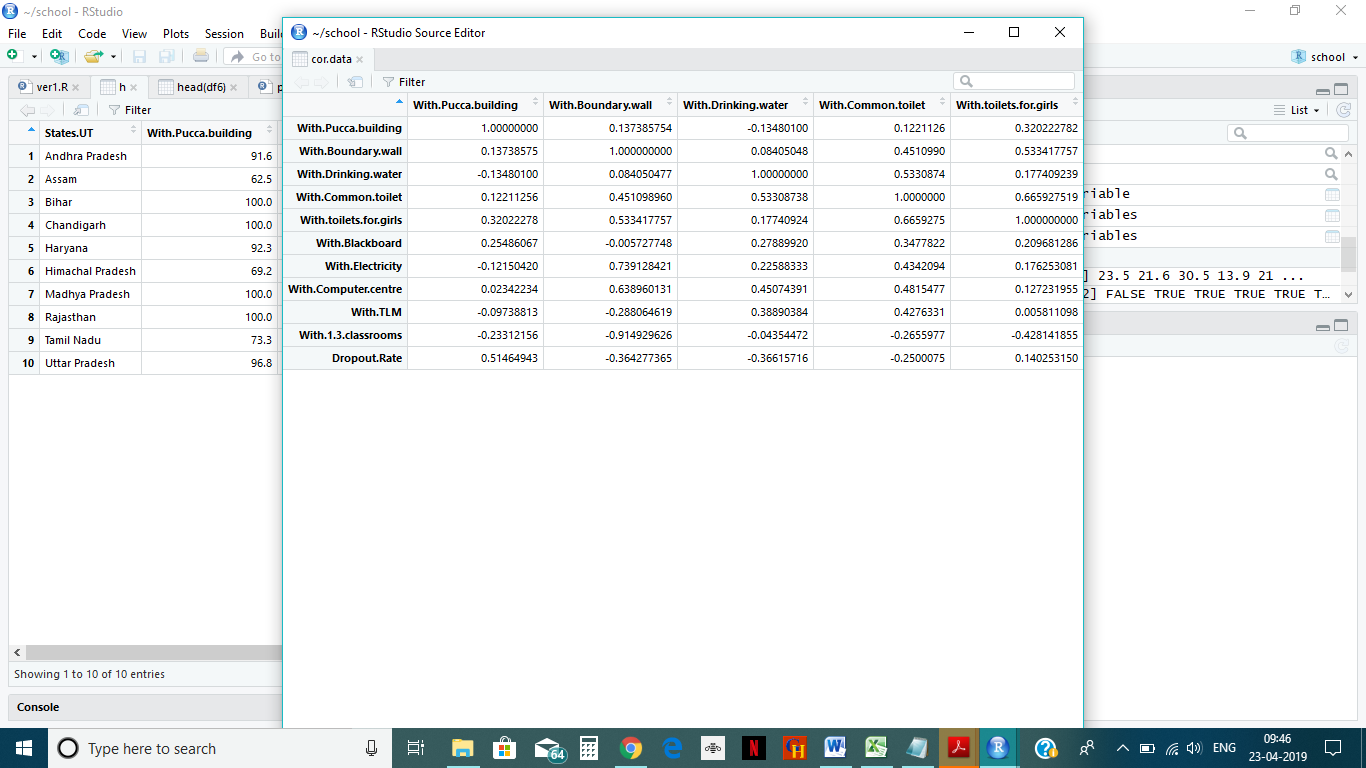


Fig 2: Corelation Matrix

After the correlation matrix, the corrgram function was used to plot the above the above correlations like so:

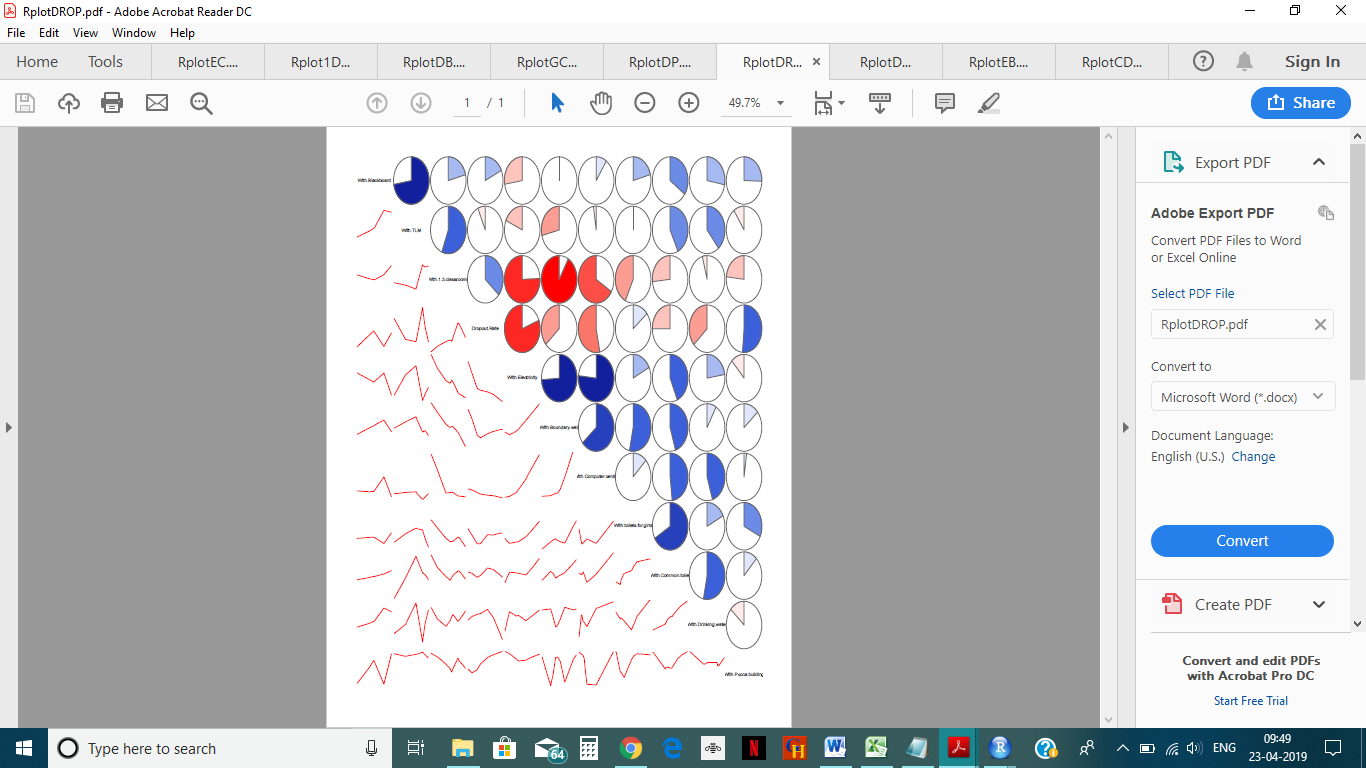


Fig 3: Corrgram function plot

Another type of graph was also plotted which clearly showed the correlations:

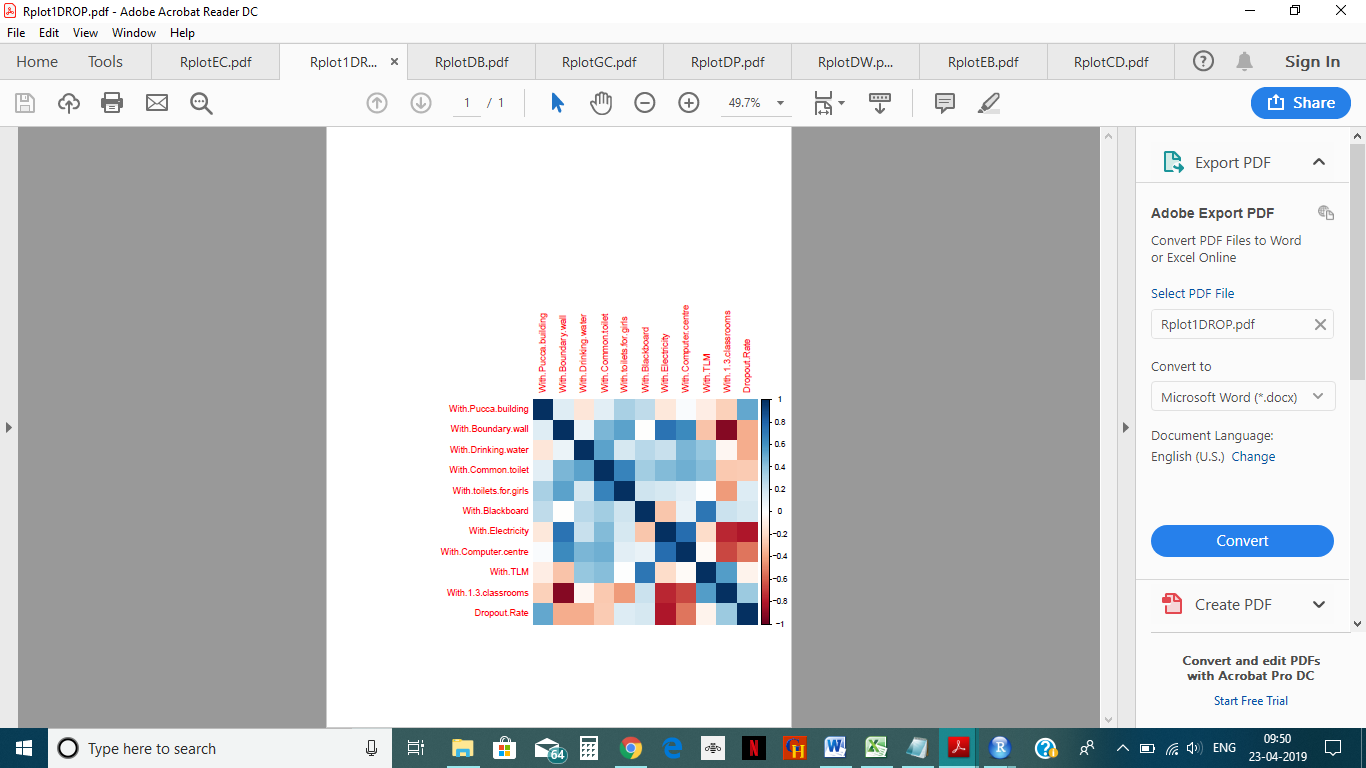
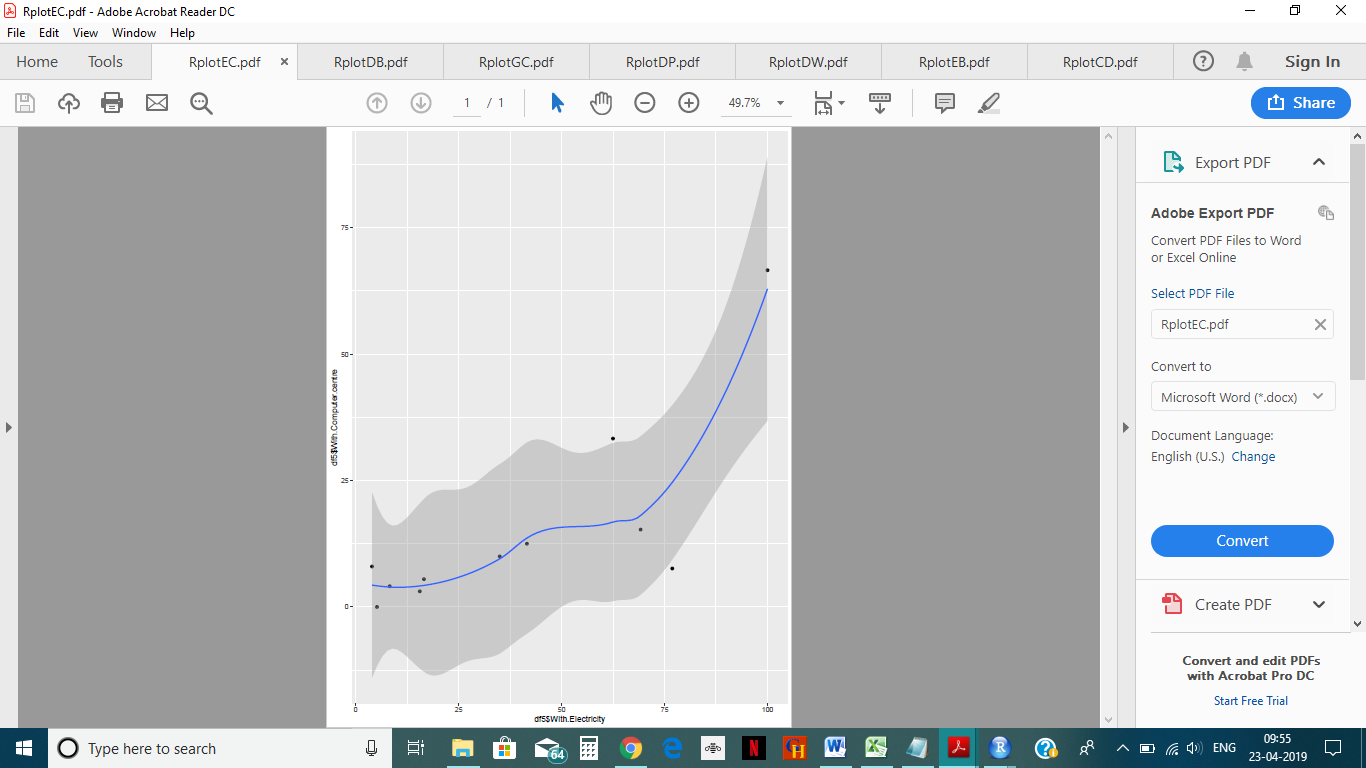
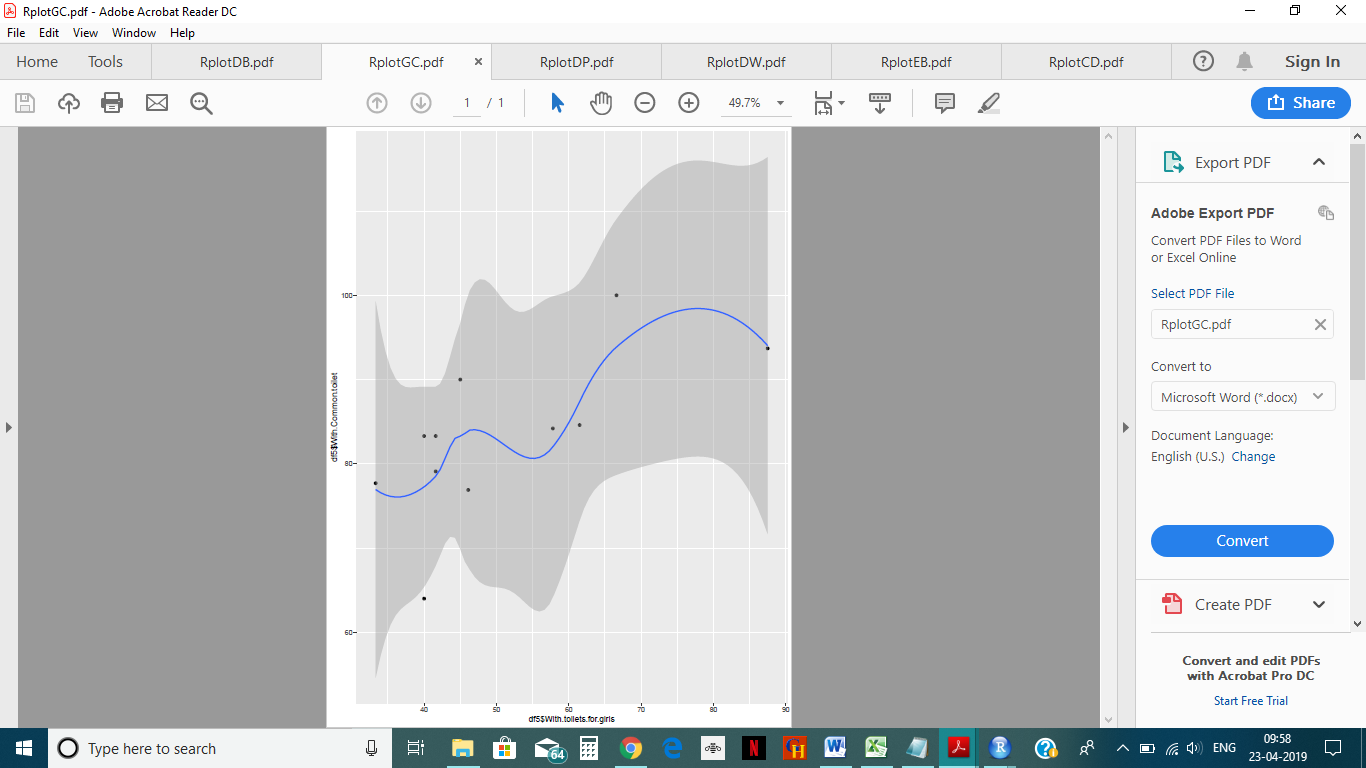


Fig 4: graph showing correlation

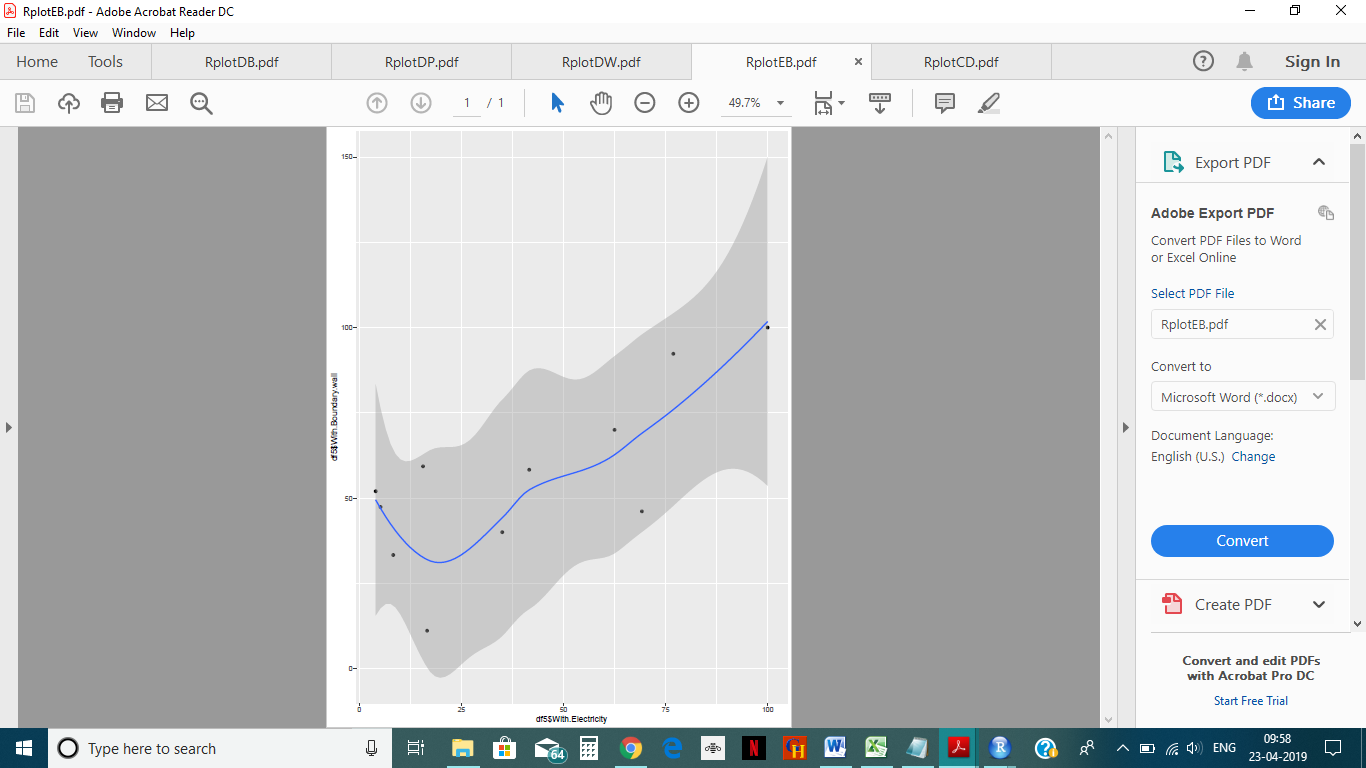
After analyzing the above graphs, the most positively correlated attributes were plotted against each other using the libraries under ggplot2. The graphs obtained were so:

Plot 1: Between schools with electricity and computer center

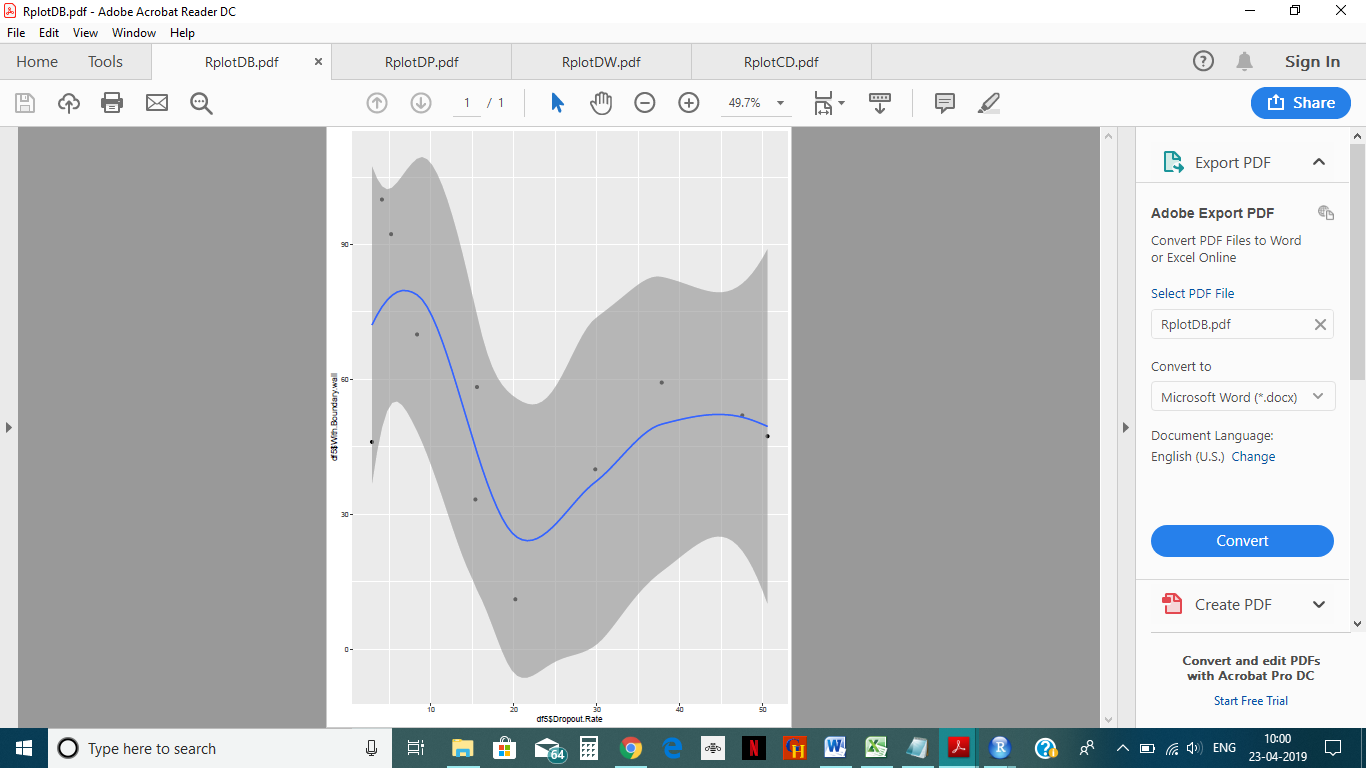
Plot 2: Between schools with girls toilet and common toilet

.

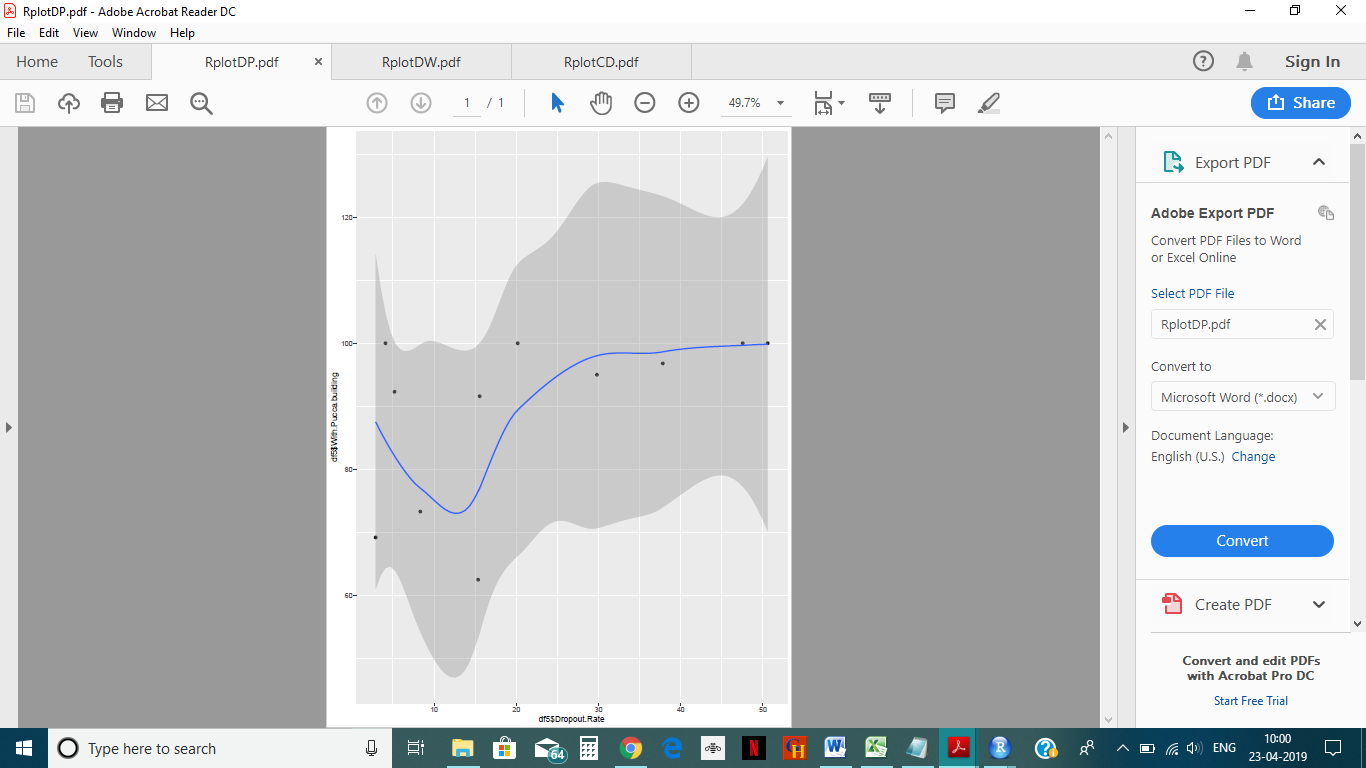
Plot 3: Between schools with electricity and boundary wall.



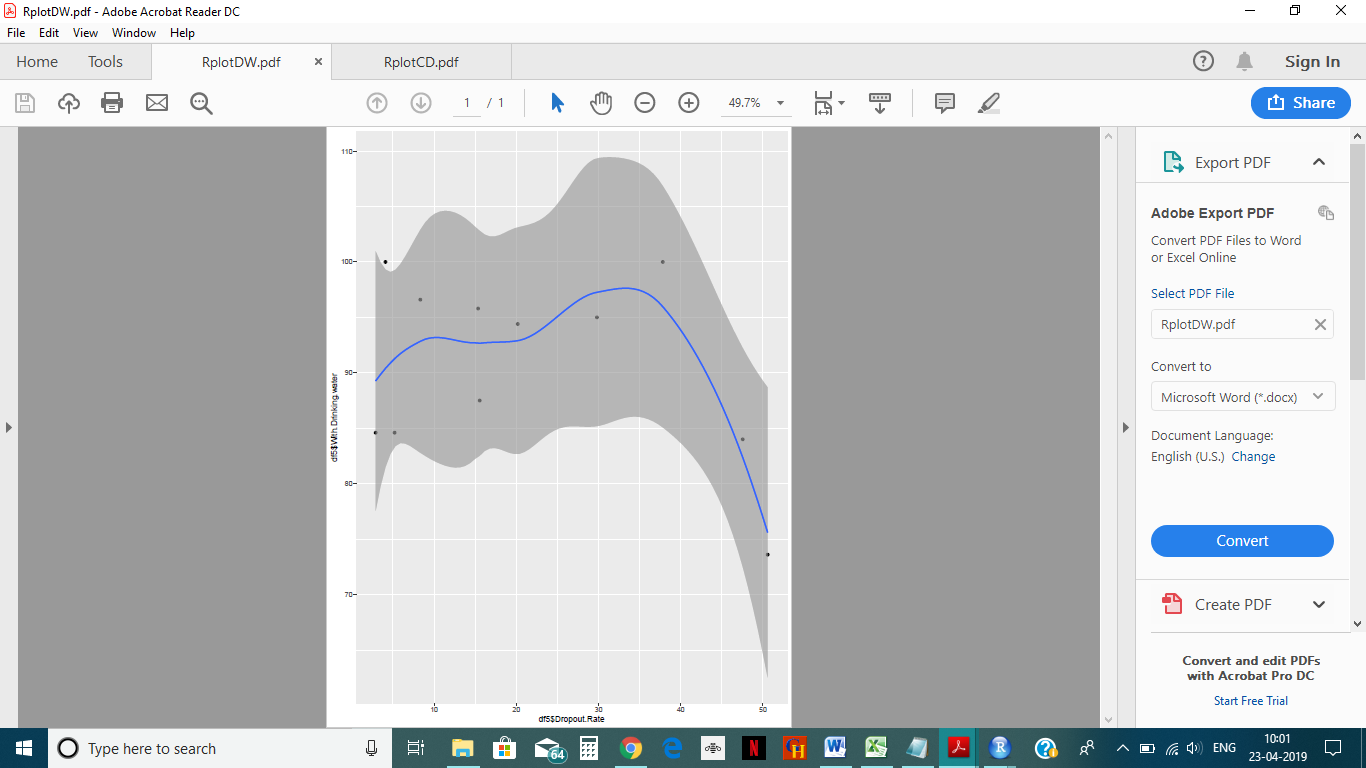
Plot 4: Between dropout rate and boundary wall.



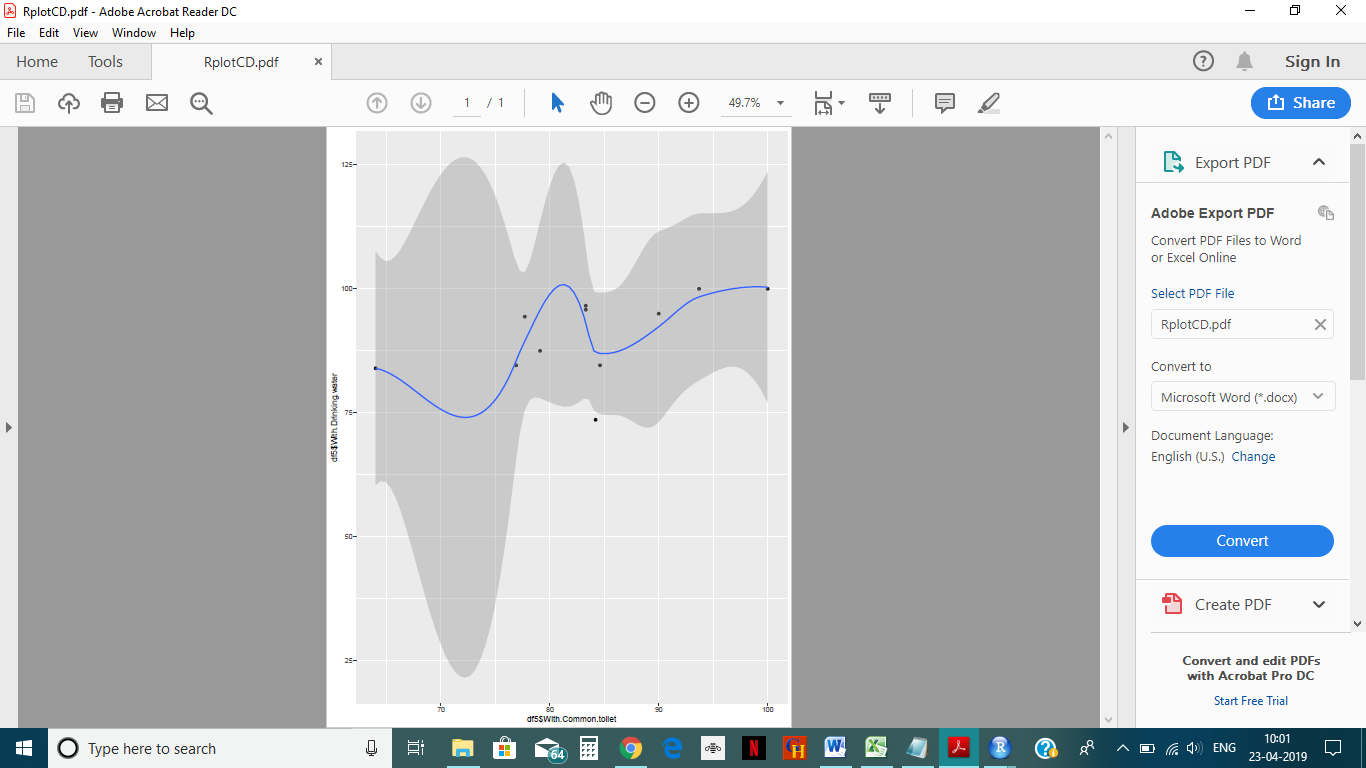
Plot 5: Between droupout rate and pucca building.



Plot 6: Between droupout rate and drinking water.



Plot 7: Between droupout rate and common toilet.



After the analysis of the above graphs and plots a model was made using liner regression to train the machine on basis of existing data and then further make predictions. The model was as follows:

#check any NA value- Clean Data

any(is.na(df6))

# Set a random see so your "random" results are the same as this notebook

set.seed(101)

str(df6)

# Split up the sample, basically randomly assigns a booleans to a new column "sample"

#sample <- sample.split(df$age, SplitRatio = 0.70) # SplitRatio = percent of sample==TRUE

#nrow(df)

#sample

# Training Data

train = subset(df)

nrow(train)

# Testing Data

#test = subset(df, sample == FALSE)

#nrow(test)

model<- lm(Dropout.Rate ~ .,train)

summary(model)

However due to lack of authentic data the model did not enough tuples to train the data. Hence the model did not make any predictions

**CHAPTER 4**

**CONCLUSION**

The various conclusions drawn from the analysis of data were as follows:

1. Dropout rate is maximum in the state of Rajasthan, followed by Bihar.
2. There was high positive correlation between dropout rate and pucca building, toilet for girls and classrooms with a blackboard.
3. There was high negative correlation between dropout rate and drinking water facility and computer center.
4. The highest positive correlation was shown between electricity and boundary wall.

Dropout is a serious problem that our country faces and we must combat this issue with proper analysis and measures.

**REFERENCES**

<https://data.gov.in/>

<https://www.tutorialspoint.com/r/r_overview.htm>

<https://www.geeksforgeeks.org/simple-linear-regression-using-r/>

<https://www.r-project.org/about.html>

**PLAGIARISM REPORT**