STUDENT PERFORMANCE ANALYSIS

A REPORT BY-

Shweta Sharma-11505295
Aman Singh- 11510296

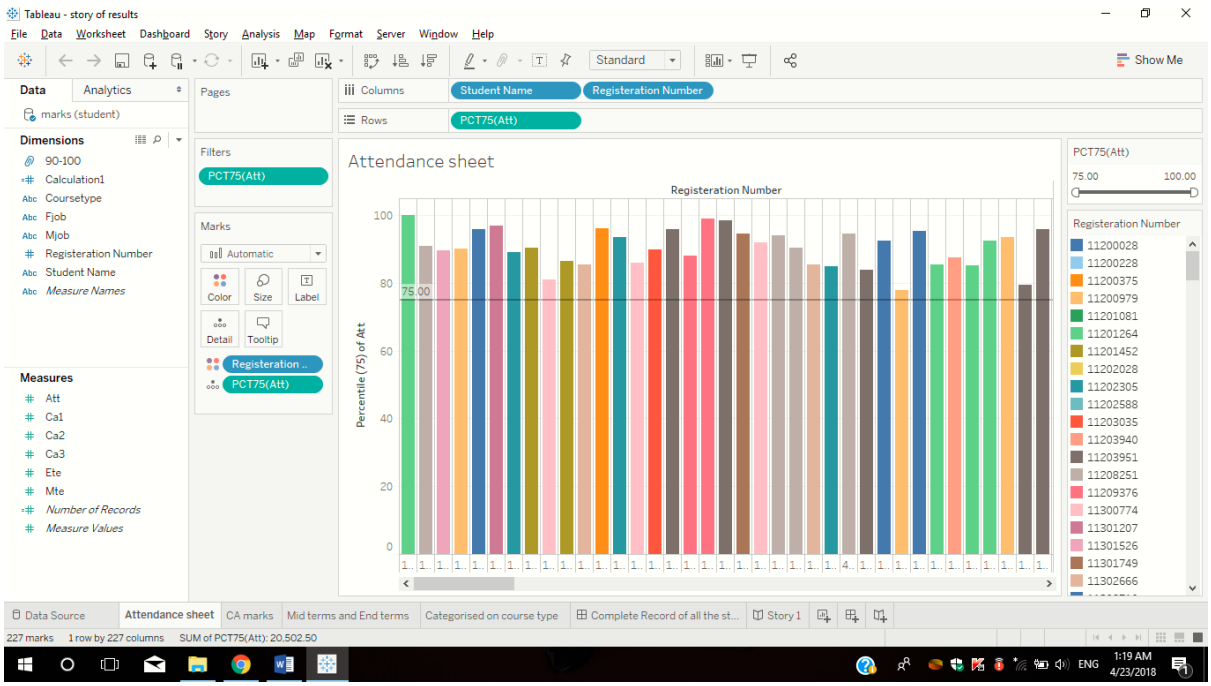
Student performance analysis

Here's the visual representation of the data:

Sheet 1: Attendance sheet

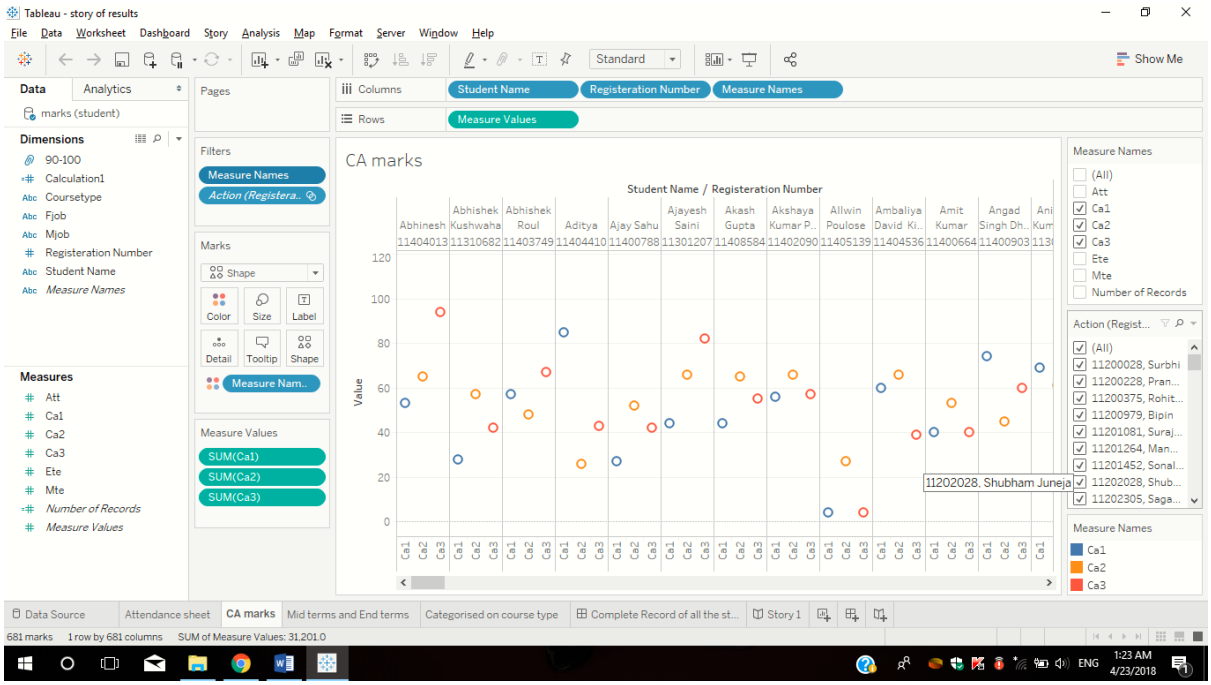
Shows the attendance of every student. The attendance is calculated as in percentage. Minimum threshold is given as 75%.

A slider filter for attendance is given to filter out students based on attendance.



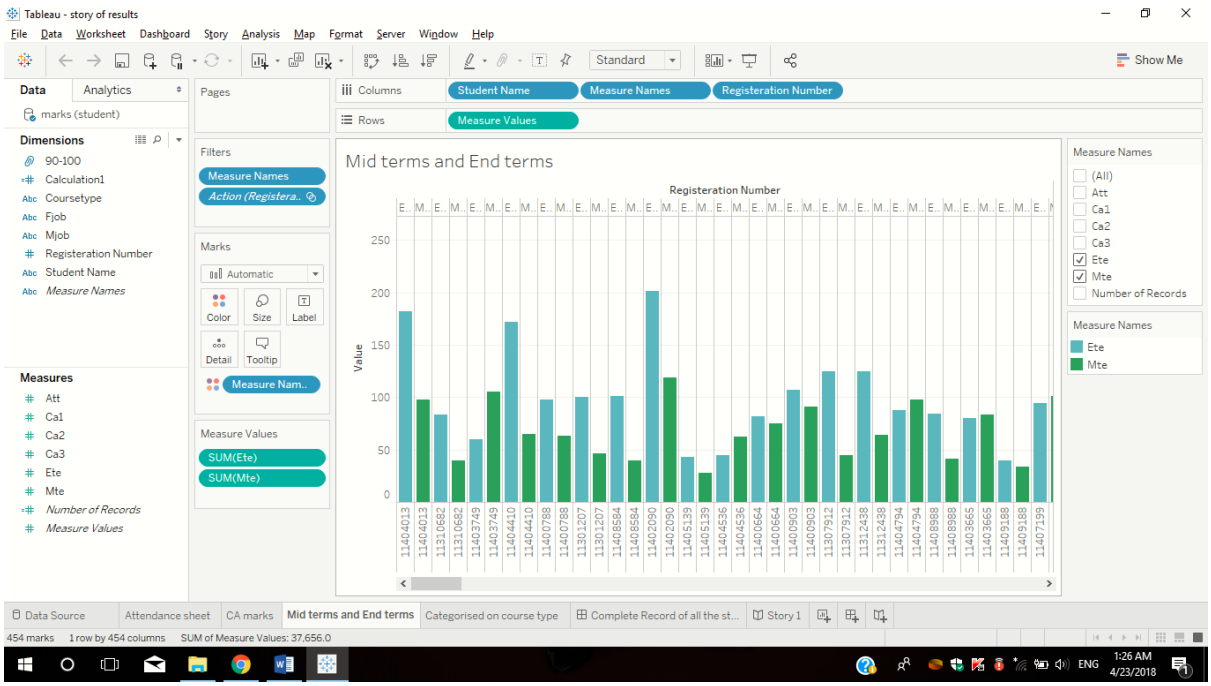
Sheet 2: CA marks

Performance of every student's all 3 CAs can be visualized from this sheet. A filter on measure names and registration numbers has been applied to this sheet.



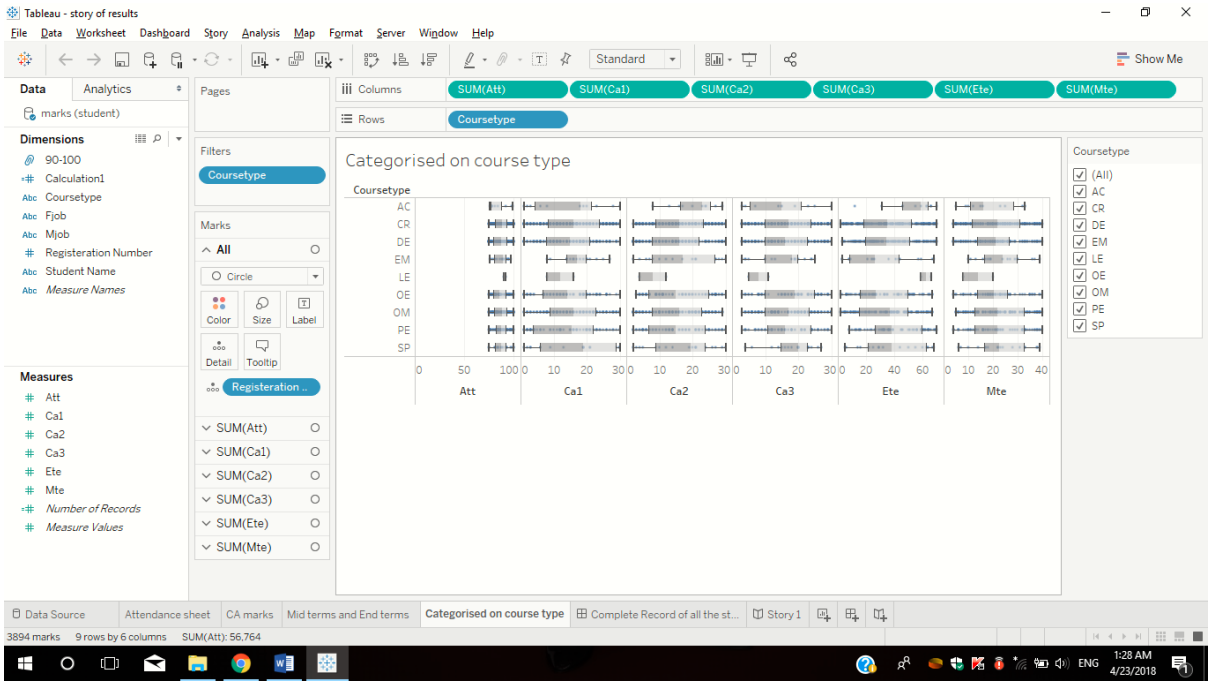
Sheet 3: Mid-terms and End terms

This sheet provides the detailed analysis and visualization of all the student's mid-term and end terms results. A filter on the measure values and registration number has been applied to the sheet.



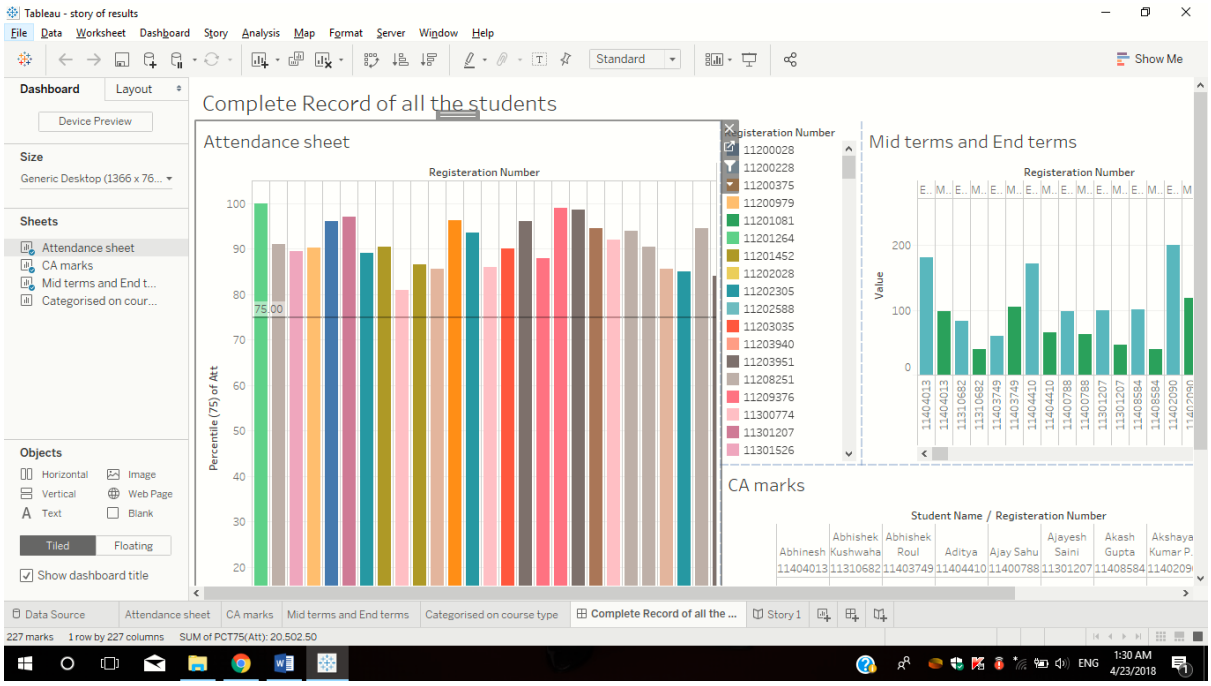
Sheet 4: categorized on course type

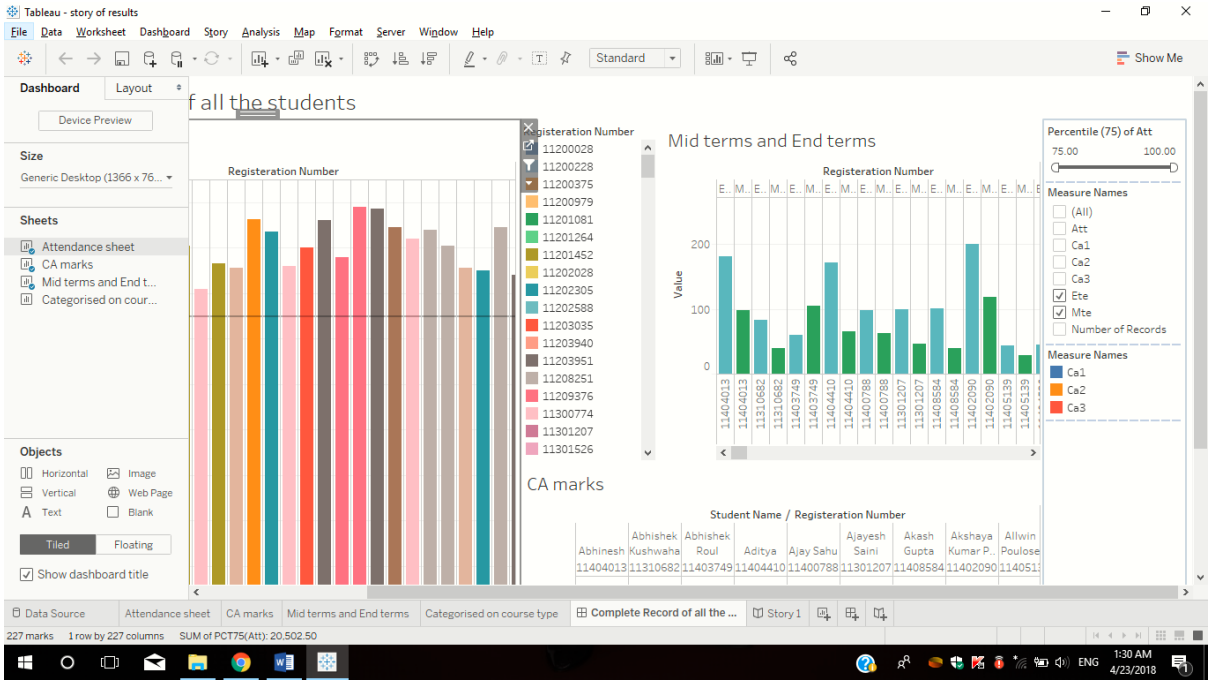
This sheet provides the detailed visualizations on all the course types. Performance of students vary based on different course types. This pattern is observed in the sheet.



Dashboard: Complete record of all the students

The dashboard provides with all the details combined together for all the students' records.





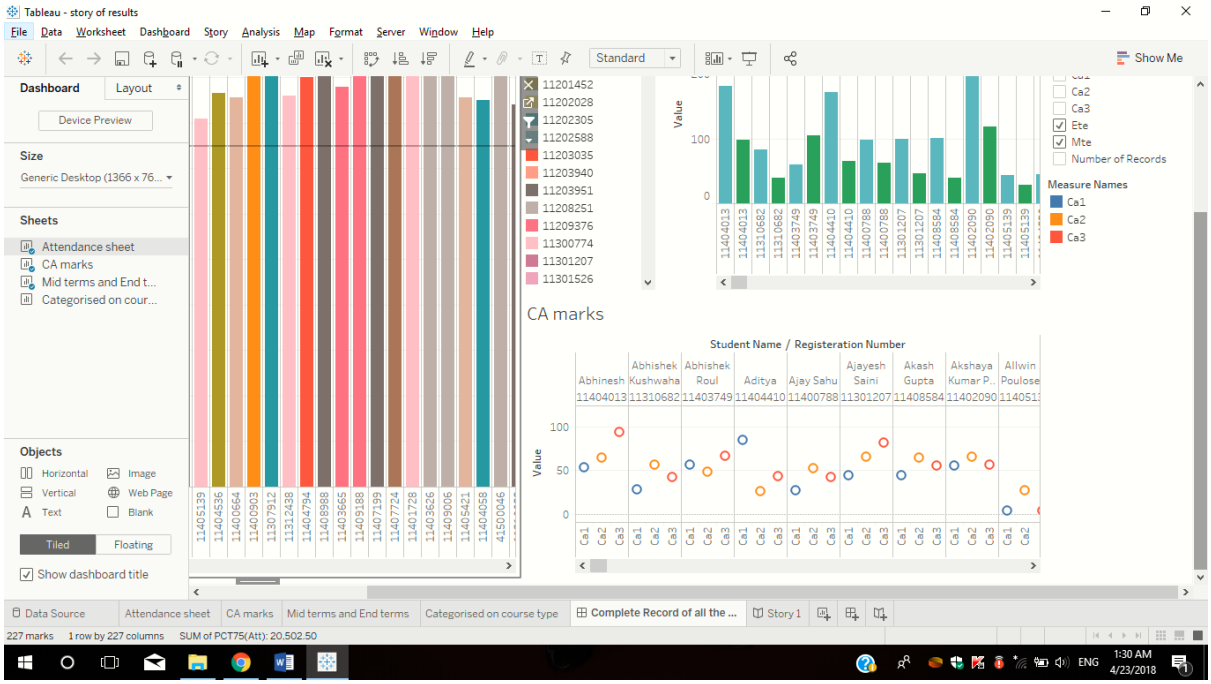


Tableau - story of results
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Dashboard
Layout
Device Preview
Size
Generic Desktop (1366 x 76...
Sheets
Attendance sheet
CA marks
Mid terms and End t...
Categorised on cour...
Objects
Horizontal
Image
Vertical
Web Page
Text
Blank
Tiled
Floating
Show dashboard title
CA marks
Student Name / Registeration Number
Abhinesh 11404013
Abhishek Kushwaha 11310682
Abhishek Roul 11403749
Aditya 11404410
Ajay Sahu 11400788
Ajayesh Saini 11301207
Akash Gupta 11408584
Akshaya Kumar P... 11402090
Allwin Poulose 1140511
Value
Measure Names
Ca1
Ca2
Ca3
Ete
Mte
Number of Records
Data Source
Attendance sheet
CA marks
Mid terms and End terms
Categorised on course type
Complete Record of all the ...
Story 1
227 marks
1 row by 227 columns
SUM of PCT75(Att): 20.502.50
ENG
1:30 AM
4/23/2018

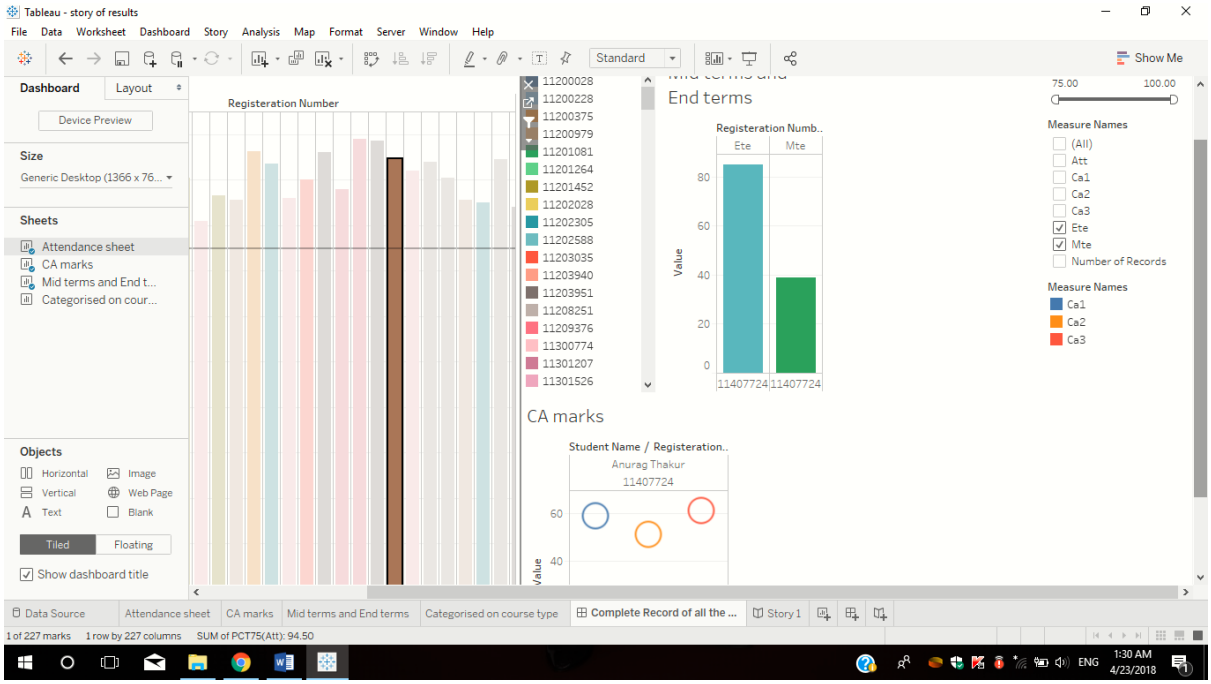
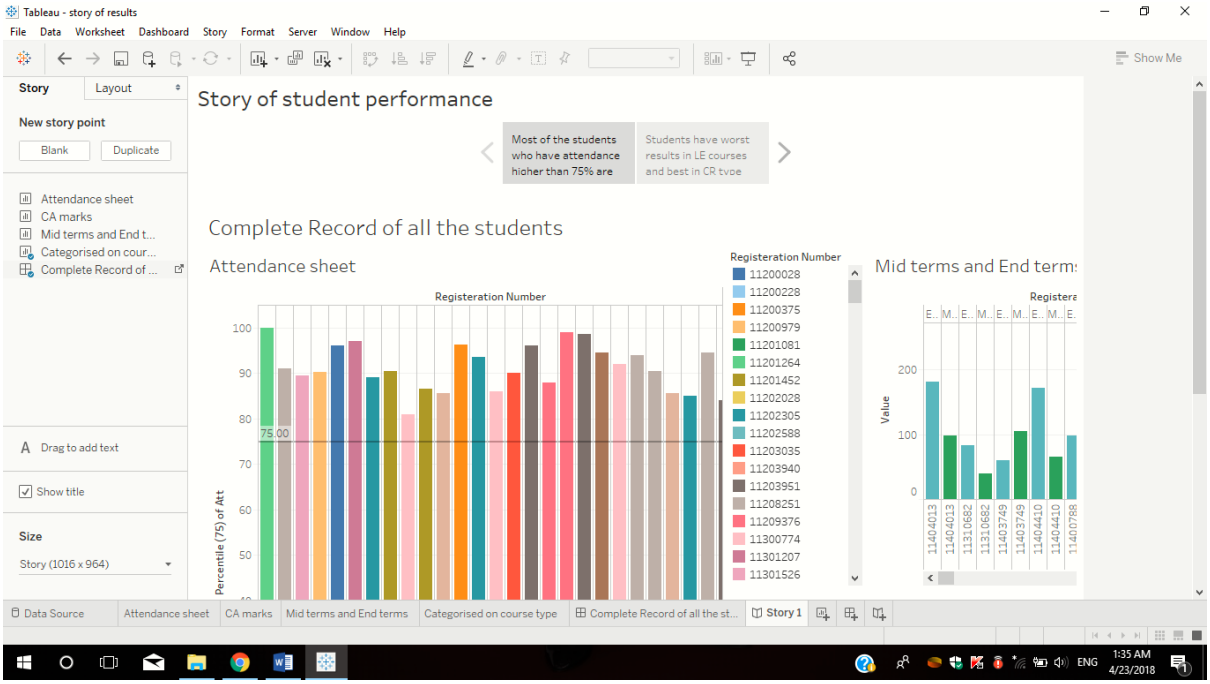


Tableau - story of results
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Dashboard
Layout
Device Preview
Size
Generic Desktop (1366 x 76...
Sheets
Attendance sheet
CA marks
Mid terms and End t...
Categorised on cour...
Objects
Horizontal
Image
Vertical
Web Page
Text
Blank
Tiled
Floating
Show dashboard title
Registeration Number
End terms
Registeration Numb..
Ete
Mte
Value
75.00
100.00
Measure Names
(All)
Att
Ca1
Ca2
Ca3
Ete
Mte
Number of Records
CA marks
Student Name / Registeration..
Anurag Thakur
11407724
Data Source
Attendance sheet
CA marks
Mid terms and End terms
Categorised on course type
Complete Record of all the ...
Story 1
1 of 227 marks
1 row by 227 columns
SUM of PCT75(Att): 94.50
ENG
1:30 AM
4/23/2018

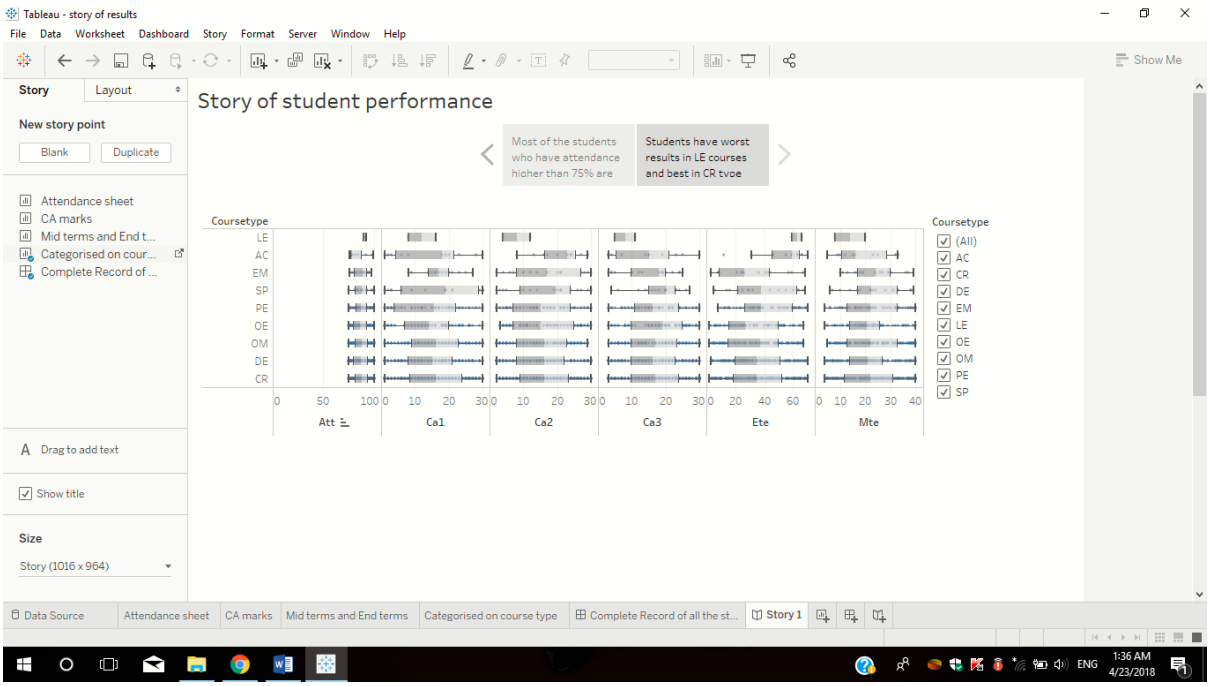
Story: Story of student performance

The story analyses 2 main viewpoints or insights from the given sheets and the dashboards:

1) Most of the students who have attendance higher than 75% are tend to perform above average in the CAs, MTEs and ETEs too.

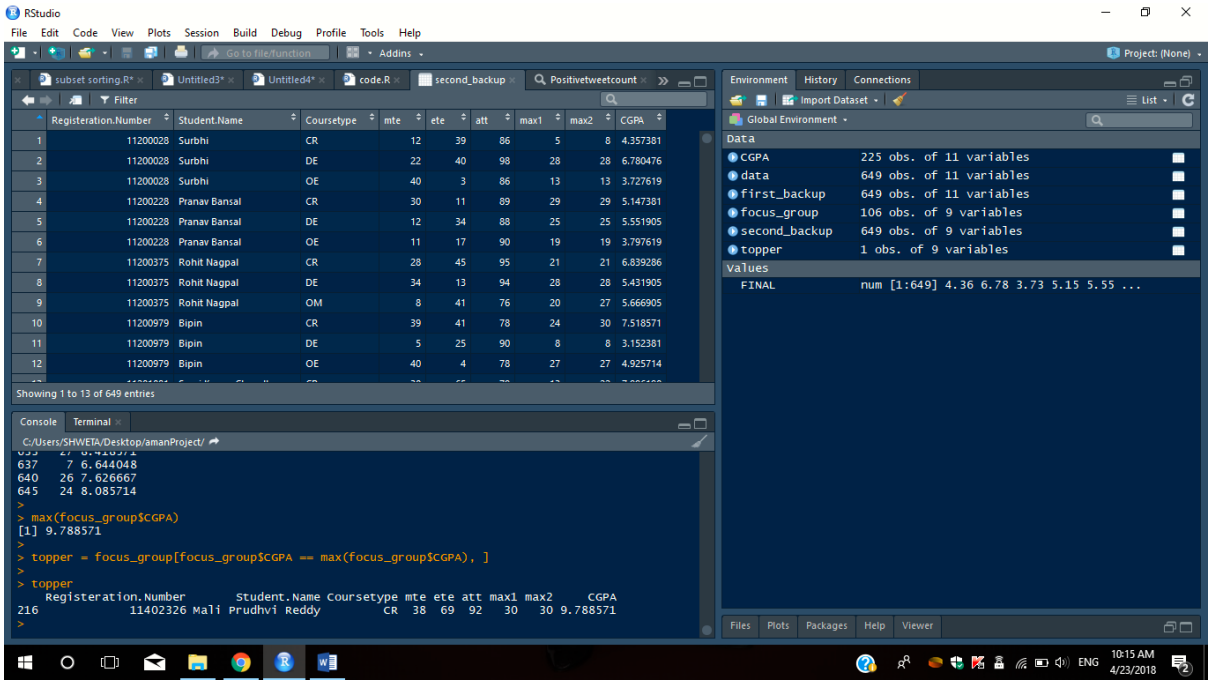


2) Students have worst results in LE courses and best in CR type courses. This depicts the interests of the students.

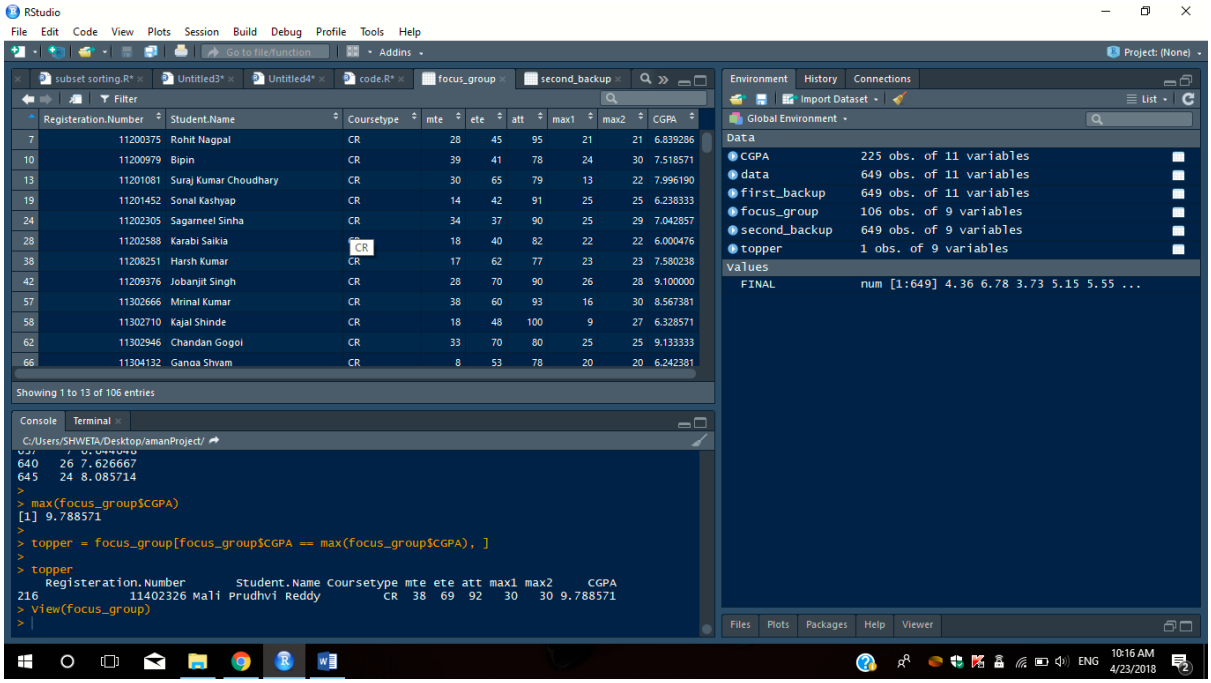


Analysis within R

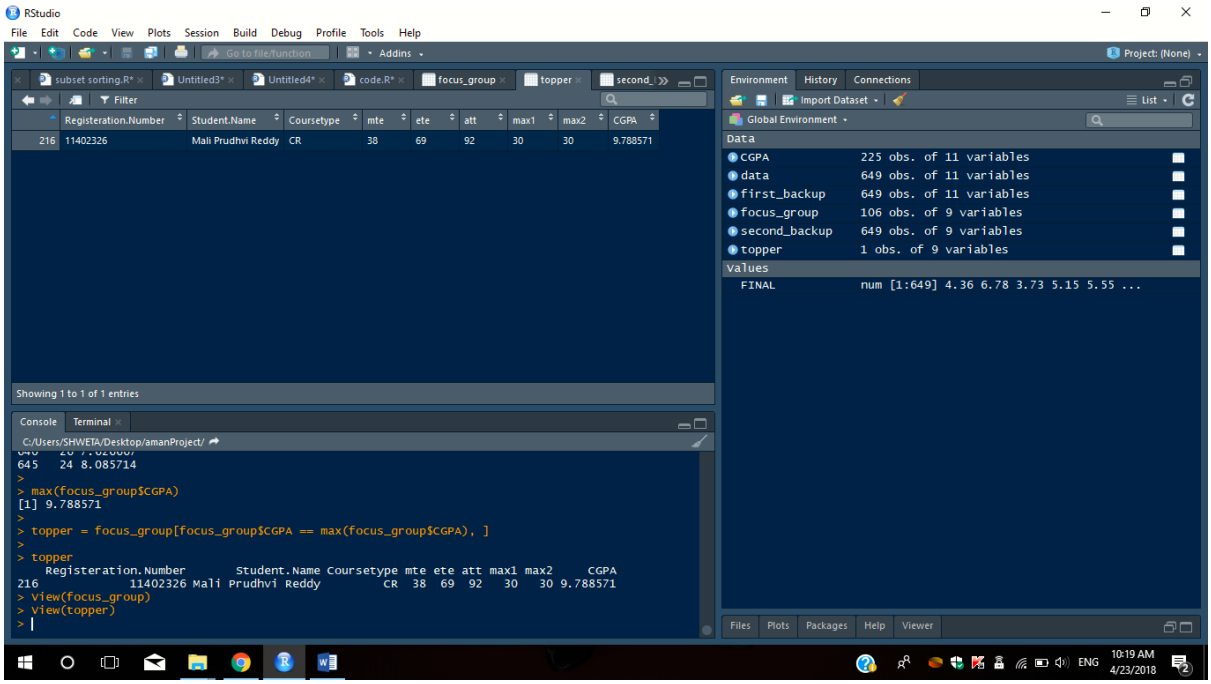
Calculated CGPA of each and every student.



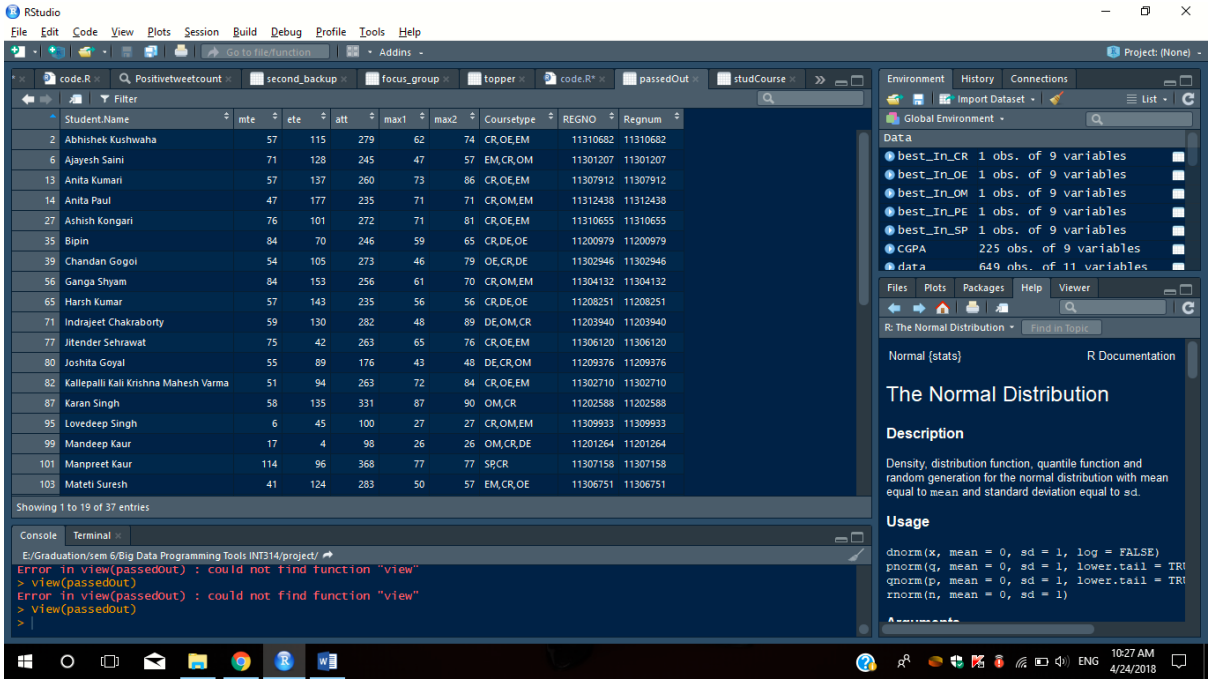
Analysed which students are eligible for being in focus group with an eligibility criteria where the students have cgpa greater than 6 in CR course type. Also found the number of students in focus group.

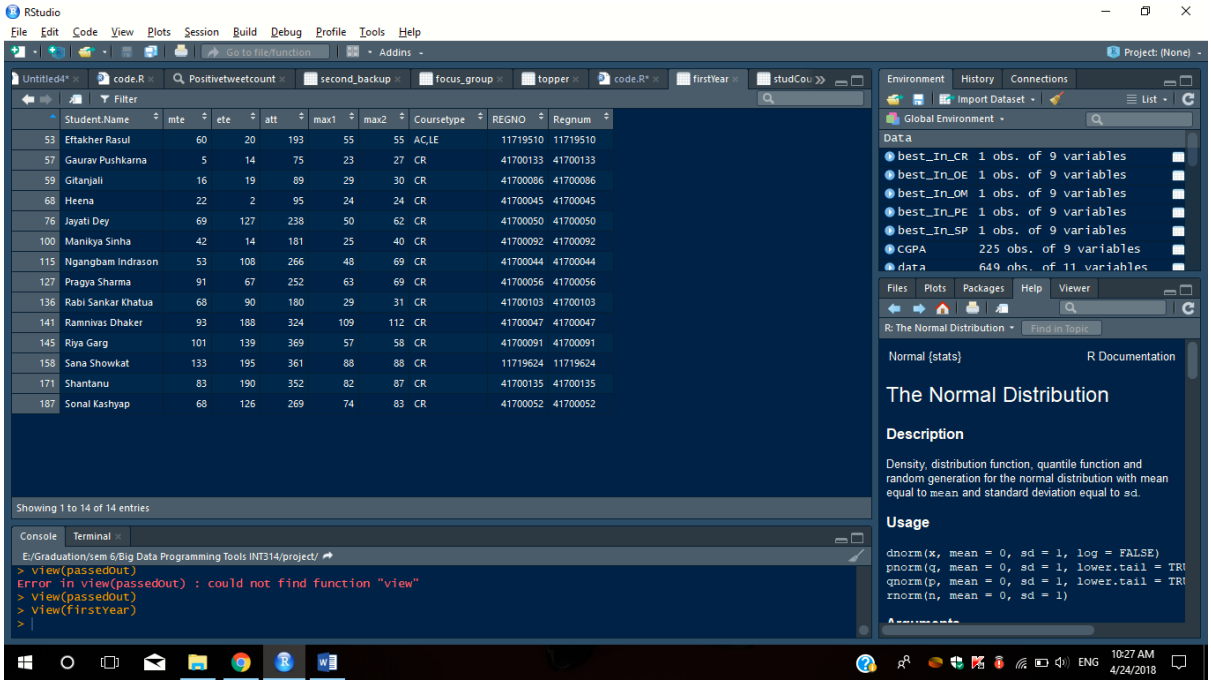


Found the topper of the class based on CR course type.

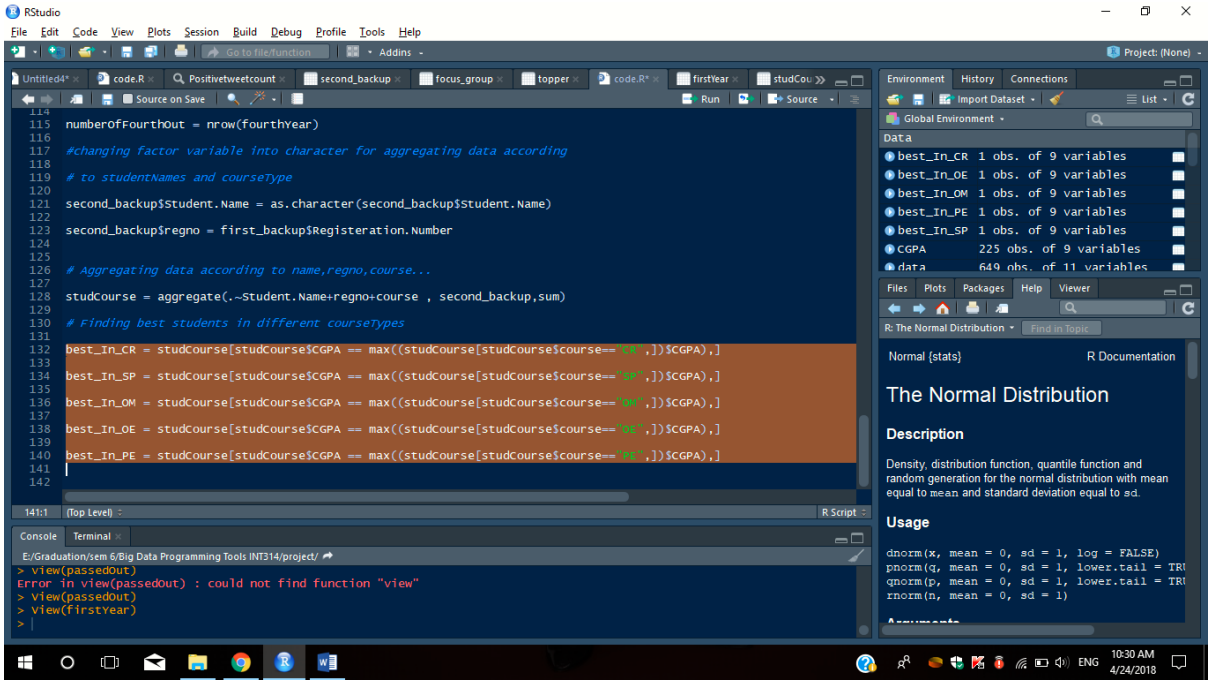


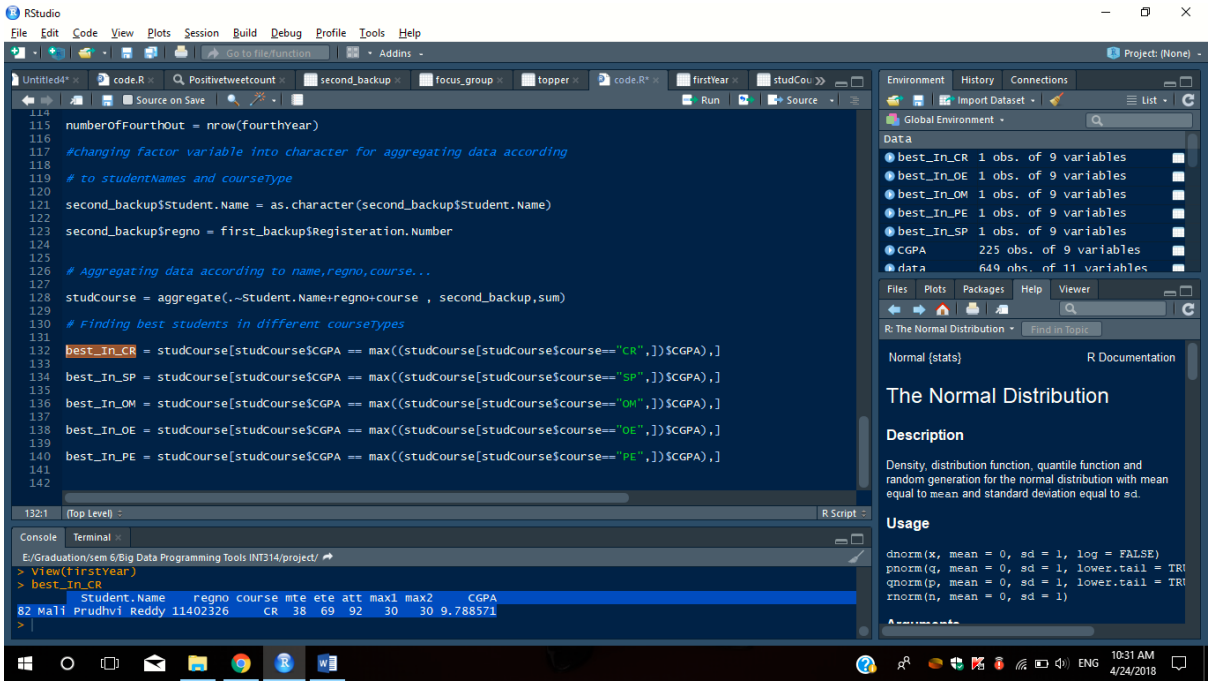
Found out how many no. of students are there in first year, second year, third year, fourth year and how many students out of the record are passed out form the college.





Finally found the topper in each course type.


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
numberOfFourthOut = nrow(fourthYear)
#changing factor variable into character for aggregating data according
to studentNames and courseType
second_backup$Student.Name = as.character(second_backup$Student.Name)
second_backup$regno = first_backup$Registeration.Number
# Aggregating data according to name,regno,course...
studCourse = aggregate(.~Student.Name+regno+course , second_backup,sum)
# Finding best students in different courseTypes
best_In_CR = studCourse[studCourse$CGPA == max((studCourse[studCourse$course== "CR",])$CGPA),]
best_In_SP = studCourse[studCourse$CGPA == max((studCourse[studCourse$course== "SP",])$CGPA),]
best_In_OM = studCourse[studCourse$CGPA == max((studCourse[studCourse$course== "OM",])$CGPA),]
best_In_OE = studCourse[studCourse$CGPA == max((studCourse[studCourse$course== "OE",])$CGPA),]
best_In_PE = studCourse[studCourse$CGPA == max((studCourse[studCourse$course== "PE",])$CGPA),]
E:/Graduation/sem 6/Big Data Programming Tools INT314/project/
> view(passedOut)
Error in view(passedOut) : could not find function "view"
> View(passedOut)
> View(firstYear)
best_In_CR 1 obs. of 9 variables
best_In_OE 1 obs. of 9 variables
best_In_OM 1 obs. of 9 variables
best_In_PE 1 obs. of 9 variables
best_In_SP 1 obs. of 9 variables
CGPA 225 obs. of 9 variables
The Normal Distribution
Description
Density, distribution function, quantile function and random generation for the normal distribution with mean equal to mean and standard deviation equal to sd.
Usage
dnorm(x, mean = 0, sd = 1, log = FALSE)
pnorm(q, mean = 0, sd = 1, lower.tail = TRU
qnorm(p, mean = 0, sd = 1, lower.tail = TRU
rnorm(n, mean = 0, sd = 1)
10:30 AM
4/24/2018


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
numberOfFourthOut = nrow(fourthYear)
#changing factor variable into character for aggregating data according
to studentNames and courseType
second_backup$Student.Name = as.character(second_backup$Student.Name)
second_backup$regno = first_backup$Registeration.Number
# Aggregating data according to name,regno,course...
studCourse = aggregate(.~Student.Name+regno+course , second_backup,sum)
# Finding best students in different courseTypes
best_In_CR = studCourse[studCourse$CGPA == max((studCourse[studCourse$course=="CR",])$CGPA),]
best_In_SP = studCourse[studCourse$CGPA == max((studCourse[studCourse$course=="SP",])$CGPA),]
best_In_OM = studCourse[studCourse$CGPA == max((studCourse[studCourse$course=="OM",])$CGPA),]
best_In_OE = studCourse[studCourse$CGPA == max((studCourse[studCourse$course=="OE",])$CGPA),]
best_In_PE = studCourse[studCourse$CGPA == max((studCourse[studCourse$course=="PE",])$CGPA),]
E:/Graduation/sem 6/Big Data Programming Tools INT314/project/
> View(firstYear)
> best_In_CR
Student.Name regno course mte ete att max1 max2 CGPA
82 Mali Prudhvi Reddy 11402326 CR 38 69 92 30 30 9.788571
best_In_CR 1 obs. of 9 variables
best_In_OE 1 obs. of 9 variables
best_In_OM 1 obs. of 9 variables
best_In_PE 1 obs. of 9 variables
best_In_SP 1 obs. of 9 variables
CGPA 225 obs. of 9 variables
The Normal Distribution
Description
Density, distribution function, quantile function and random generation for the normal distribution with mean equal to mean and standard deviation equal to sd.
Usage
dnorm(x, mean = 0, sd = 1, log = FALSE)
pnorm(q, mean = 0, sd = 1, lower.tail = TRU
qnorm(p, mean = 0, sd = 1, lower.tail = TRU
rnorm(n, mean = 0, sd = 1)
10:31 AM
4/24/2018

Here goes the code for the R analysis:

```
library(plyr)

getwd()

setwd("D://project//big data//project//project")

data <- read.csv("dataSet.csv")

first_backup <- data

second_backup = first_backup

#max of 2 ca

second_backup$max1 <- pmax(first_backup$ca1,first_backup$ca2)

second_backup$max2 <- pmax(second_backup$max1,first_backup$ca3)

second_backup$ca1=NULL

second_backup$ca2=NULL

second_backup$ca3=NULL
```

```
second_backup$Mjob=NULL

second_backup$Fjob=NULL

# aggregating data of same row

newData <- ddply(second_backup, .(Student.Name,Registeration.Number),
summarize,

        max1=paste(max1,collapse=","),

        max2=paste(max2,collapse=",") ,

         mte= paste(mte,collapse=","),

         ete=paste(ete,collapse=","),

         Coursetype = paste(Coursetype,collapse = ","))

Course = newData$Coursetype

REGNO = newData$Registeration.Number

# dataset for cgpa calculation

second_backup$Registeration.Number=NULL

second_backup$Coursetype=NULL
```

```
CGPA = aggregate(. ~ Student.Name, second_backup, sum)

CGPA$Coursetype = Course[1:225]

CGPA$REGNO = REGNO[1:225]

#CGPA calculation subject wise

FINAL = (second_backup$mte * (20/40) + second_backup$ete * (50/70) +
second_backup$att * (5/100) + (second_backup$max1 +
second_backup$max2) * (25/60)) / 10

# Adding cgpa in our dataset

second_backup$CGPA = FINAL

#adding courseType

second_backup$course = first_backup$Coursetype

# creating dataset for focus group students

focus_group = second_backup[second_backup$CGPA >= 6 &
second_backup$course=="CR",]

# counting number of students in focus group
```

```
numberOfStudentsInFocusGroup = nrow(focus_group)

# finding toppers data and CGPA

topper = focus_group[focus_group$CGPA == max(focus_group$CGPA), ]

print(topper)

#changing registration number(integer) into character for finding
1st,2nd,3rd,4rth year and passed out students

CGPA$Regnum = as.character(REGNO[1:225])

# Dataset for different year students

passedOut =
CGPA[substr(CGPA$Regnum,3,3)=='2'|substr(CGPA$Regnum,3,3)=='3',]
View(passedOut)

firstYear = CGPA[substr(CGPA$Regnum,3,3)=='7',]
View(firstYear)

secondYear = CGPA[substr(CGPA$Regnum,3,3)=='6',]
View(secondYear)

thirdYear = CGPA[substr(CGPA$Regnum,3,3)=='5',]
```

```
View(thirdYear)

fourthYear = CGPA[substr(CGPA$Regnum,3,3)=='4',]
View(fourthYear)

#calculating number of students in 1st,2nd,3rd,4rth year and passed out

numberOfPassedOut =  nrow(passedOut)

numberOfFirstYear =  nrow(firstYear)

numberOfSecondYear = nrow(secondYear)

numberOfThirdYear = nrow(thirdYear)

numberOfFourthOut = nrow(fourthYear)

#changing factor variable into character for aggregating data according

# to studentNames and courseType

second_backup$Student.Name = as.character(second_backup$Student.Name)

second_backup$regno = first_backup$Registeration.Number
```

```
# Aggregating data according to name,regno,course...

studCourse = aggregate(.~Student.Name+regno+course , second_backup,sum)

# Finding best students in different courseTypes

best_In_CR = studCourse[studCourse$CGPA ==
max((studCourse[studCourse$course=="CR",])$CGPA),]

best_In_SP = studCourse[studCourse$CGPA ==
max((studCourse[studCourse$course=="SP",])$CGPA),]

best_In_OM = studCourse[studCourse$CGPA ==
max((studCourse[studCourse$course=="OM",])$CGPA),]

best_In_OE = studCourse[studCourse$CGPA ==
max((studCourse[studCourse$course=="OE",])$CGPA),]

best_In_PE = studCourse[studCourse$CGPA ==
max((studCourse[studCourse$course=="PE",])$CGPA),]
```

THANK YOU