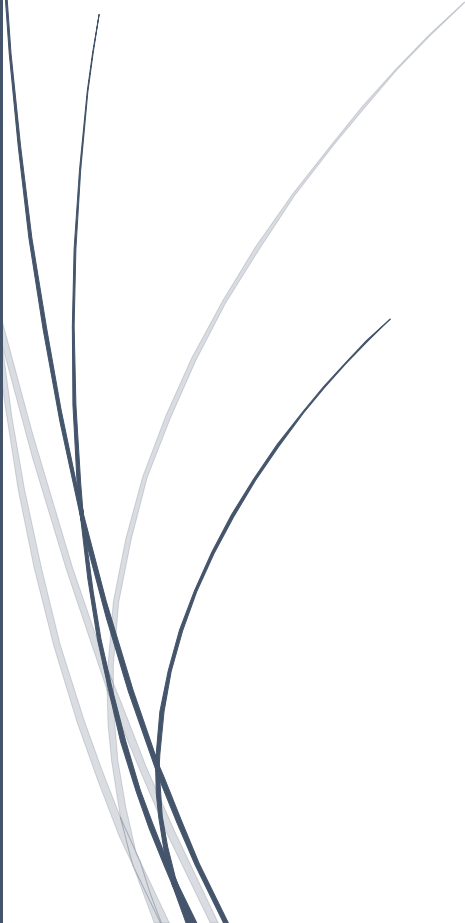


A dark blue vertical bar runs down the left side of the page. A blue arrow points to the right from this bar, containing the date.

3/23/2023

# Mathematics Project

Analysis on Dropout Rate of  
Students from School Over the  
Years

Several thin, curved lines in shades of blue and grey sweep upwards from the bottom left corner of the page.

Sakhil Taneja (43)  
Dushyant Kumar Singh (23)  
Deepak Pal (21)

<u>Sr. No.</u>	<u>Topic</u>	<u>Pg. No.</u>
1	Modifying and Filtering The Data	04
2	Analysis Based upon Different Class Groups	06
3	Gender Based Analysis in Different States	11
4	Gender Based Analysis over Different Class Groups	14
5	Normality Check for Primary School Dropout rate	18

## Dataset Description:

Title: Dropout rate of Students among different Classes of School

State_UT	year	Primary_E	Primary_C	Primary_T	Upper Pri	Upper Pri	Upper Pri	Secondary	Secondary
A & N Isl	2012-13	0.83	0.51	0.68	NR	1.09	1.23	5.57	5.55
A & N Isl	2013-14	1.35	1.06	1.21	NR	1.54	0.51	8.36	5.98
A & N Isl	2014-15	0.47	0.55	0.51	1.44	1.95	1.69	11.47	8.16
Andhra Pr	2012-13	3.3	3.05	3.18	3.21	3.51	3.36	12.21	13.25
Andhra Pr	2013-14	4.31	4.39	4.35	3.46	4.12	3.78	11.95	13.37
Andhra Pr	2014-15	6.57	6.89	6.72	5.09	5.32	5.2	15.81	15.6
Arunachal	2013-14	11.54	10.22	10.89	4.44	6.74	5.59	16.08	12.75
Arunachal	2012-13	15.84	14.44	15.16	5.86	9.06	7.47	13.99	11.77
Arunachal	2014-15	11.51	10.09	10.82	5.31	8.08	6.71	18.33	15.81
Assam	2012-13	7.02	5.46	6.24	7.89	6.55	7.2	25.65	27.79
Assam	2013-14	8.19	6.68	7.44	7.6	6.54	7.05	28.59	32.1
Assam	2014-15	16.07	14.65	15.36	10.45	10.56	10.51	24.64	29.28
Bihar	2012-13	NR	NR	NR	NR	NR	NR	29.11	31.28
Bihar	2013-14	2.38	1.79	2.09	2.77	3.19	2.98	24.67	26.05
Bihar	2014-15	0.35	NR	NR	4.14	4.01	4.08	25.21	26.62
Chandigar	2012-13	NR	NR	NR	NR	0.88	0.1	NR	NR
Chandigar	2013-14	NR	NR	NR	0.72	1.55	1.08	NR	NR
Chandigar	2014-15	NR	NR	NR	0.01	0.96	0.44	NR	NR
Chhattisga	2012-13	4.24	4.05	4.14	6.09	4.73	5.42	14.82	14.89
Chhattisga	2013-14	1.45	1.38	1.42	4.09	3.51	3.8	24.1	22.72

Source: data.gov.in

Links to our data: <https://data.gov.in/resource/drop-out-rate-2012-13-2014-15>

**Brief Into to the Dataset:** This is a Data Set of students dropping out of School among different Stated from years 2012-13 to 2014-15,

Classes are divided into four Class groups:

- Primary
- Upper Primary
- Secondary
- HrSecondary(Higher Secondary)

And each group is divided into Three Parts:

- Boys
- Girls
- Total

Importing Our Dataset into RStudio:

```
mydata1<-read.csv("DOR.csv")
```

Here mydata1 is a variable which contains our data

## 1) Modifying the datatype and Filtering the data based on Year:

#changing the datatype of all the columns of 'mydata1' except the first two.

#Since our data is of type 'Character' in columns with NA(empty) observation so we need to convert it into datatype 'Numeric'

```
mydata1 <- mutate_at(mydata1, vars(Primary_Boys, Primary_Girls,
Primary_Total, `Upper_Primary_Boys`, `Upper_Primary_Girls`, `Upper_Primary_Total`,
`Secondary_Boys`, `Secondary_Girls`, `Secondary_Total`, HrSecondary_Boys,
HrSecondary_Girls, HrSecondary_Total),
as.numeric)
```

#extracting the data of different sessions from 'mydata1'

```
year1213<-mydata1 %>%
  filter(year == '2012-13')
year1314<-mydata1 %>%
  filter(year == '2013-14')
year1415<-mydata1 %>%
  filter(year == '2014-15')
```

Here variable year1213 contains our data for year 2012-13, Similarly year1314 & year1415 contains data for year 2013-14 & 2014-15 simultaneously for our dataset mydata1

**Data set for year 12-13**

State_UT	year	Primary_E	Primary_C	Primary_T	Upper.Pri	Upper.Pri	Upper.Pri	Secondary	Secondary
A & N Isl	2012-13	0.83	0.51	0.68	NR	1.09	1.23	5.57	5.55
Andhra Pr	2012-13	3.3	3.05	3.18	3.21	3.51	3.36	12.21	13.25
Arunachal	2012-13	15.84	14.44	15.16	5.86	9.06	7.47	13.99	11.77
Assam	2012-13	7.02	5.46	6.24	7.89	6.55	7.2	25.65	27.79
Bihar	2012-13	NR	NR	NR	NR	NR	NR	29.11	31.28
Chandigar	2012-13	NR	NR	NR	NR	0.88	0.1	NR	NR
Chhattisga	2012-13	4.24	4.05	4.14	6.09	4.73	5.42	14.82	14.89
Dadra & N	2012-13	NR	0.47	NR	2.59	4.3	3.38	26.34	22
Daman & D	2012-13	NR	NR	NR	NR	NR	NR	10.84	12.25
Delhi	2012-13	NR	NR	NR	NR	0.31	NR	6.26	4.21
Goa	2012-13	NR	0.26	NR	NR	NR	NR	9.49	6.26
Gujarat	2012-13	0.21	1.35	0.74	2.75	8.19	5.2	13.96	12.95
Haryana	2012-13	1.48	1.06	1.29	0.18	0.8	0.46	8.22	9.93
Himachal	2012-13	0.51	0.39	0.45	0.52	0.49	0.51	8.35	8.4
Jammu & C	2012-13	6.8	5.75	6.3	5.51	5.52	5.52	16.79	17.96
Jharkhand	2012-13	7.36	7.05	7.21	4.99	5.94	5.47	17.84	19.2
Karnataka	2012-13	3.4	2.51	2.97	4.96	5.15	5.05	40.7	39.07
Kerala	2012-13	NR	NR	NR	NR	NR	NR	12.31	6.38

**Data set for year 13-14**

State_UT	year	Primary_E	Primary_C	Primary_T	Upper.Pri	Upper.Pri	Upper.Pri	Secondary	Secondary	Secondary
A & N Isl	2013-14	1.35	1.06	1.21	NR	1.54	0.51	8.36	5.98	7.2
Andhra Pr	2013-14	4.31	4.39	4.35	3.46	4.12	3.78	11.95	13.37	12.65
Arunachal	2013-14	11.54	10.22	10.89	4.44	6.74	5.59	16.08	12.75	14.49
Assam	2013-14	8.19	6.68	7.44	7.6	6.54	7.05	28.59	32.1	30.43
Bihar	2013-14	2.38	1.79	2.09	2.77	3.19	2.98	24.67	26.05	25.33
Chandigar	2013-14	NR	NR	NR	0.72	1.55	1.08	NR	NR	NR
Chhattisga	2013-14	1.45	1.38	1.42	4.09	3.51	3.8	24.1	22.72	23.41
Dadra & N	2013-14	1.05	2.17	1.57	3.31	5.09	4.13	30.02	26.83	28.59
Daman & D	2013-14	1.06	0.59	0.85	3.42	3.72	3.56	19.46	20.33	19.86
Delhi	2013-14	NR	NR	NR	3.13	2.37	2.78	10.8	6.64	8.9
Goa	2013-14	0.08	0.33	0.2	NR	NR	NR	11.68	7.19	9.58
Gujarat	2013-14	0.5	1.06	0.76	3.52	8.04	5.55	22.85	19.81	21.61
Haryana	2013-14	0.22	0.63	0.41	1.97	3.25	2.55	11.92	13.27	12.51
Himachal	2013-14	0.57	0.34	0.46	0.6	0.98	0.78	9.32	8.83	9.09
Jammu & C	2013-14	5.53	5.37	5.46	3.86	4.8	4.3	14.7	16.14	15.36

**Data set for year 14-15**

State_UT	year	Primary_E	Primary_G	Primary_T	Upper.Pri	Upper.Pri	Upper.Pri	Secondary	Secondary
A & N Isl	2014-15	0.47	0.55	0.51	1.44	1.95	1.69	11.47	8.16
Andhra Pr	2014-15	6.57	6.89	6.72	5.09	5.32	5.2	15.81	15.6
Arunachal	2014-15	11.51	10.09	10.82	5.31	8.08	6.71	18.33	15.81
Assam	2014-15	16.07	14.65	15.36	10.45	10.56	10.51	24.64	29.28
Bihar	2014-15	0.35	NR	NR	4.14	4.01	4.08	25.21	26.62
Chandigar	2014-15	NR	NR	NR	0.01	0.96	0.44	NR	NR
Chhattisga	2014-15	3.08	2.74	2.91	6.47	5.22	5.85	22.62	19.92
Dadra & N	2014-15	1.6	1.31	1.47	3.7	4.4	4.02	18.19	15.08
Daman & I	2014-15	1.8	0.29	1.11	3.14	3.07	3.11	34.45	29.73
Delhi	2014-15	NR	NR	NR	0.95	0.55	0.76	13.55	9.8
Goa	2014-15	0.63	0.84	0.73	NR	0.21	0.07	14.15	7.68
Gujarat	2014-15	0.82	0.98	0.89	4.65	8.54	6.41	26.29	23.24
Haryana	2014-15	5.54	5.69	5.61	5.5	6.18	5.81	16.35	15.3
Himachal	2014-15	0.46	0.83	0.64	0.5	1.27	0.87	6.31	5.8
Jammu &	2014-15	6.98	6.59	6.79	4.98	5.95	5.44	16.97	17.65
Jharkhand	2014-15	5.91	5.03	5.48	9.01	8.96	8.99	23.45	24.56

## 2) Plotting the data to differentiate b/w dropout rate of States based upon different Class groups:

**->plotting the bargraph for the dropout rate of PRIMARY CLASS students**

```
compareable_data = rbind(year1213[,5],year1314[,5],year1415[,5] )
colnames(compareable_data) = year1213[,1]
rownames(compareable_data) = c("2012-13", "2013-14","2014-15")
barplot(compareable_data, beside=TRUE, legend=TRUE, cex.names = 0.5, las =
3)
```

**->plotting the bargraph for the dropout rate of UPPER PRIMARY CLASS students**

```
compareable_data2 = rbind(year1213[,8],year1314[,8],year1415[,8] )
colnames(compareable_data2) = year1213[,1]
rownames(compareable_data2) = c("2012-13", "2013-14","2014-15")
barplot(compareable_data2, beside=TRUE, legend=TRUE, cex.names = 0.5, las
= 3)
```

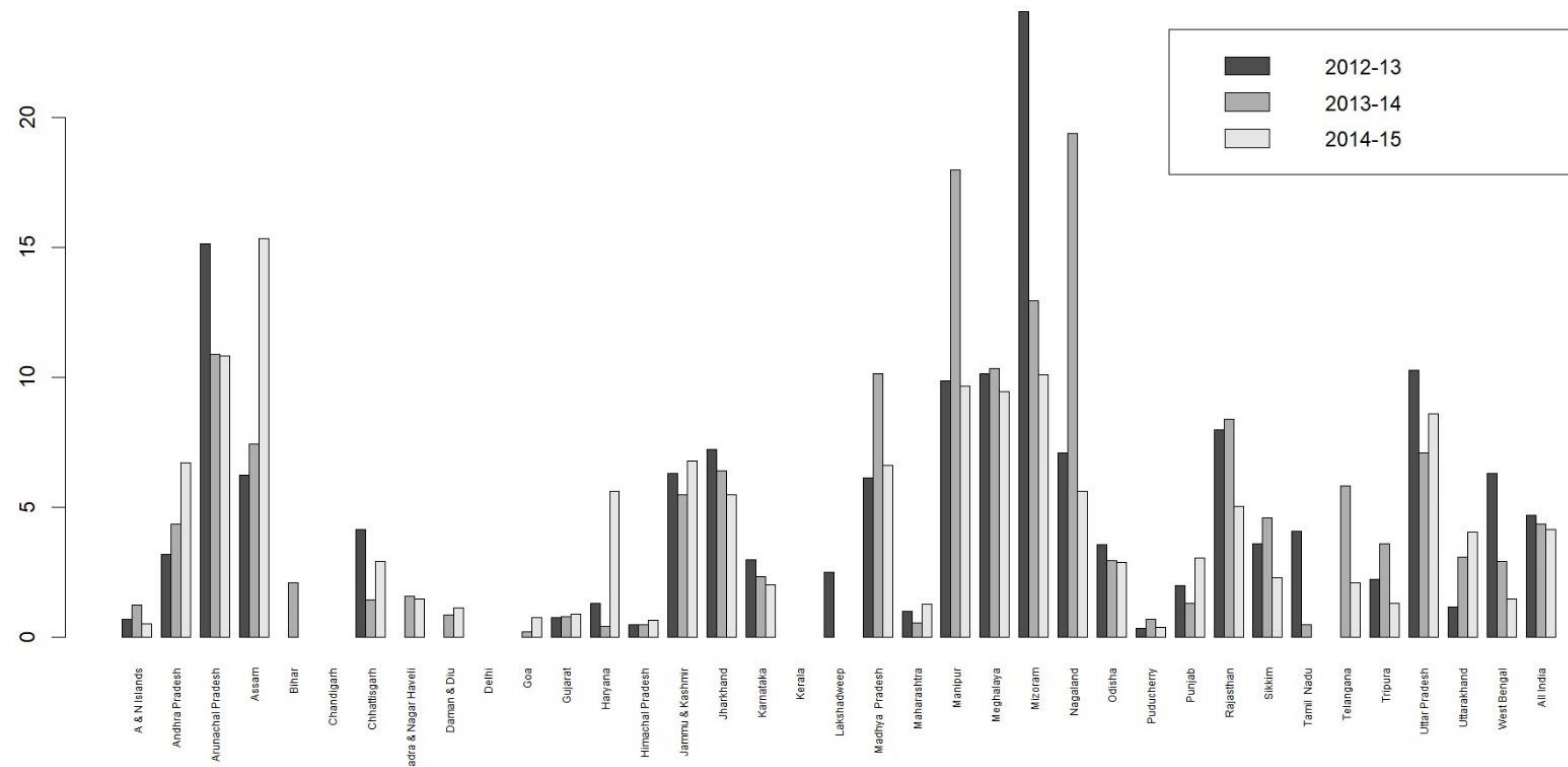
->plotting the bargraph for the dropout rate of **SECONDARY CLASS** students

```
compareable_data3 = rbind(year1213[,11],year1314[,11],year1415[,11] )
colnames(compareable_data3) = year1213[,1]
rownames(compareable_data3) = c("2012-13", "2013-14","2014-15")
barplot(compareable_data3, beside=TRUE, legend=TRUE, cex.names = 0.5, las
= 3)
```

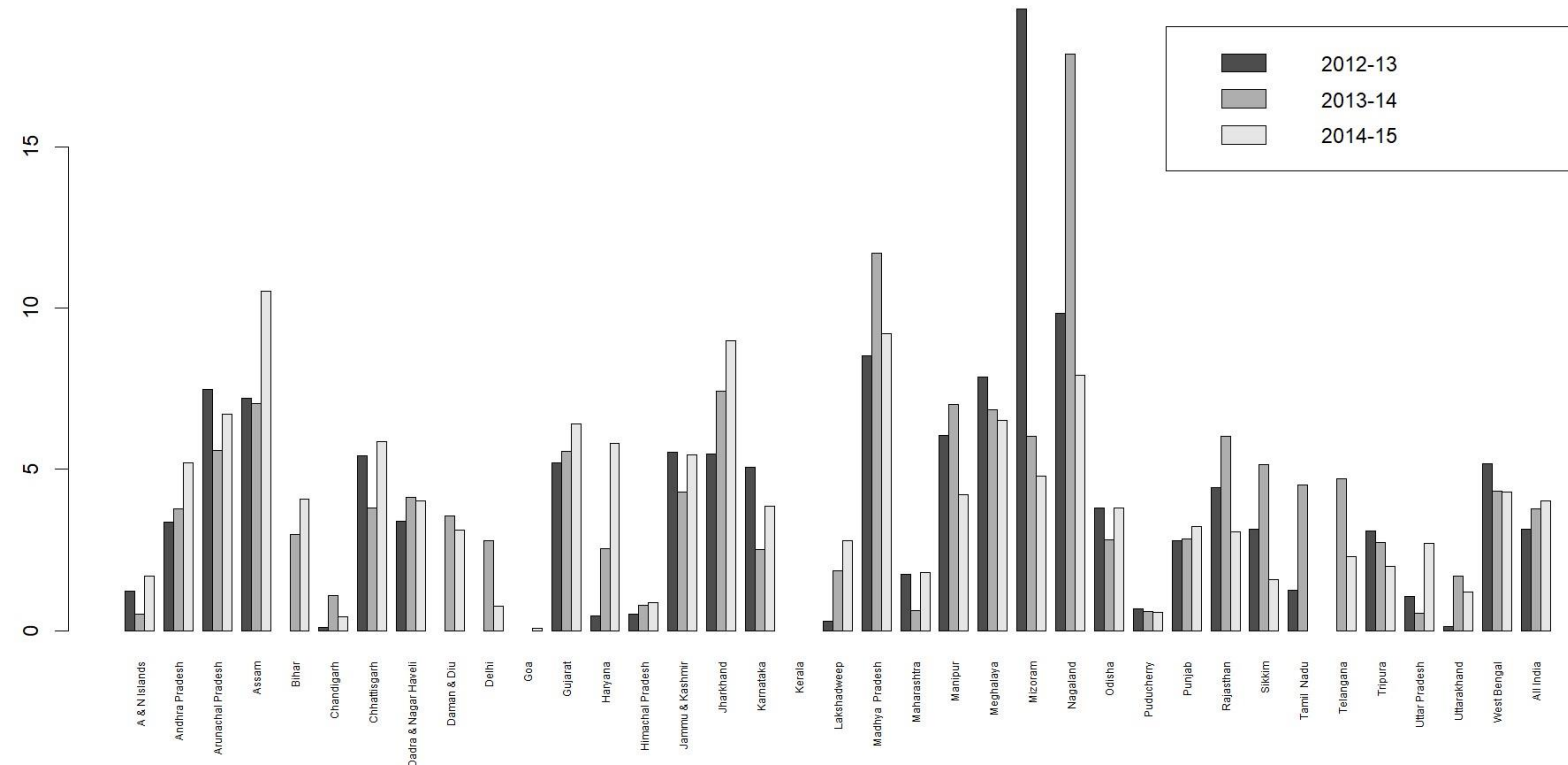
->plotting the bargraph for the dropout rate of **HIGHER SCEONDARY CLASS** students

```
compareable_data4 = rbind(year1213[,14],year1314[,14],year1415[,14] )
colnames(compareable_data4) = year1213[,1]
rownames(compareable_data4) = c("2012-13", "2013-14","2014-15")
barplot(compareable_data4, beside=TRUE, legend=TRUE, cex.names = 0.5, las
= 3)
```

**Primary Total**

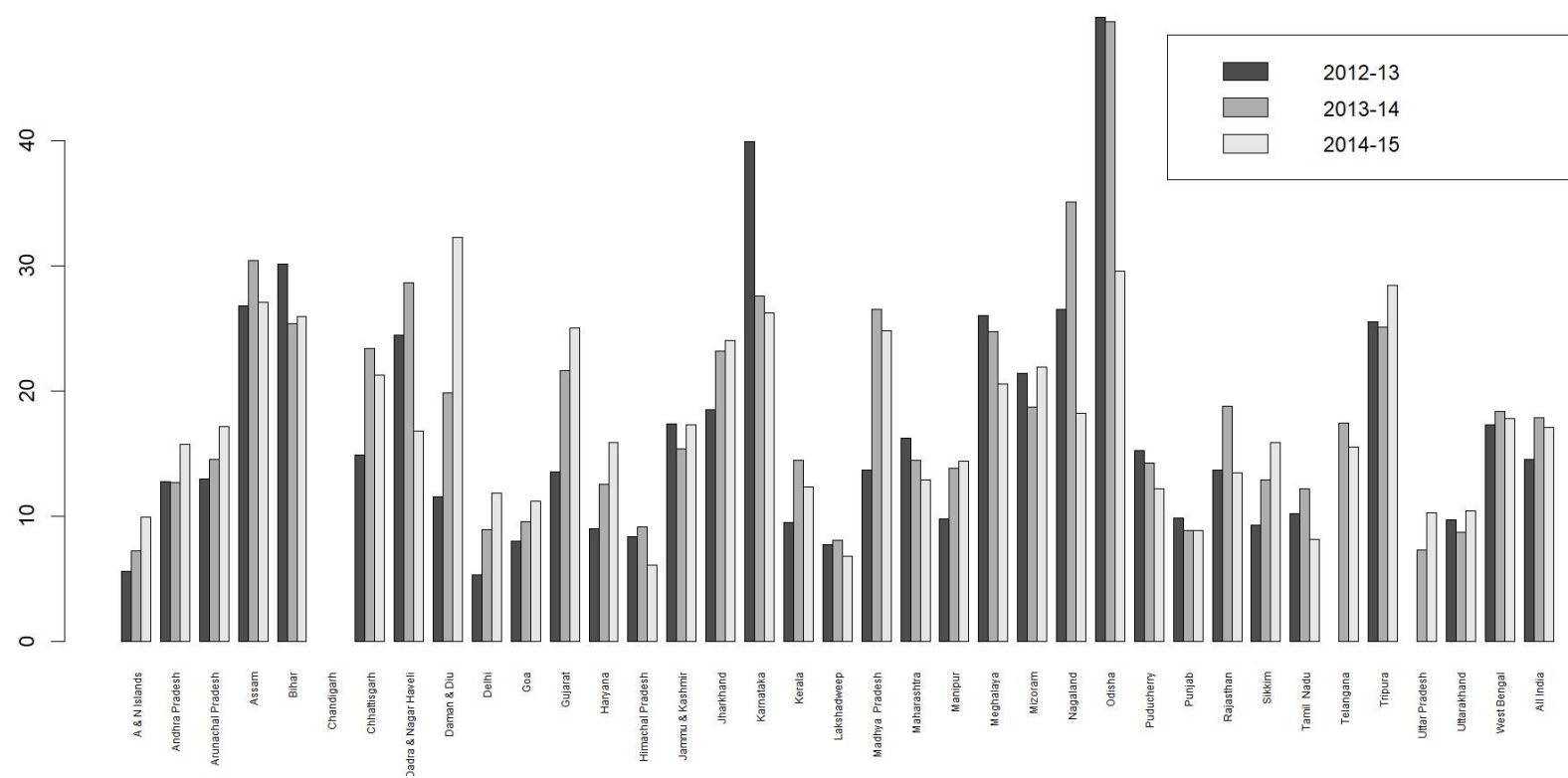


## Upper Primary Total

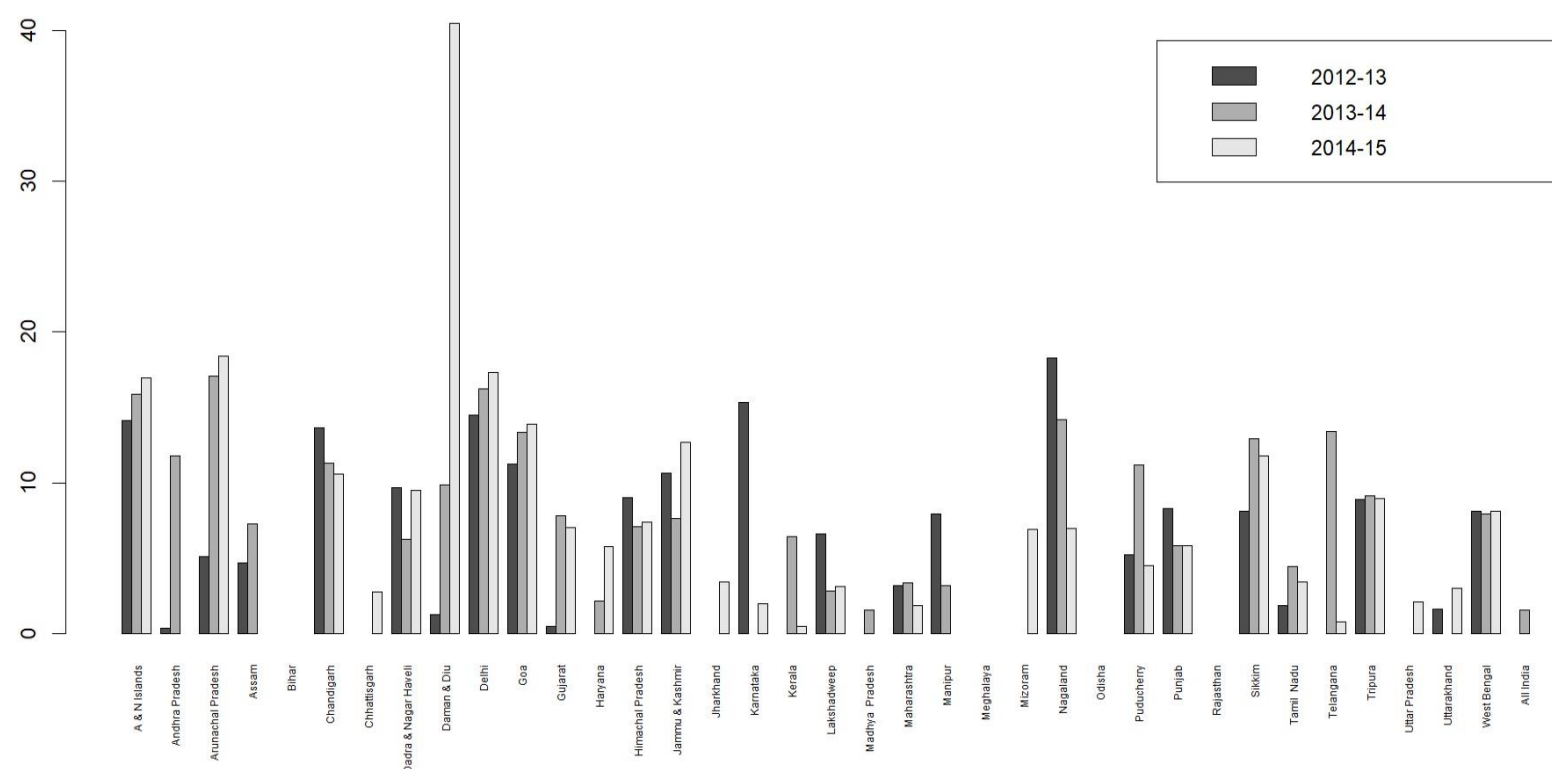




## Secondary Total



## Hr Secondary Total



Now that we have data of Comparison of Drop Out rates in States we can observe that:

- For Primary School: Mizoram (2012-13) has Highest Dropout rate
- For Upper Primary School: Mizoram (2012-13) has the Highest Dropout rate
- For Secondary School: Odisha(2012-13) has the Highest Dropout rate with Odisha(2013-14) being 2<sup>nd</sup> close
- For Higher Secondary School: Daman and Diu (2014-15) has the Highest Dropout Rate

Since Students dropping School in Primary School is a really big Concern Let us focus at that for now:

3)Plotting the data to Differentiate b/w dropout rates of Boys and Girls in different States:

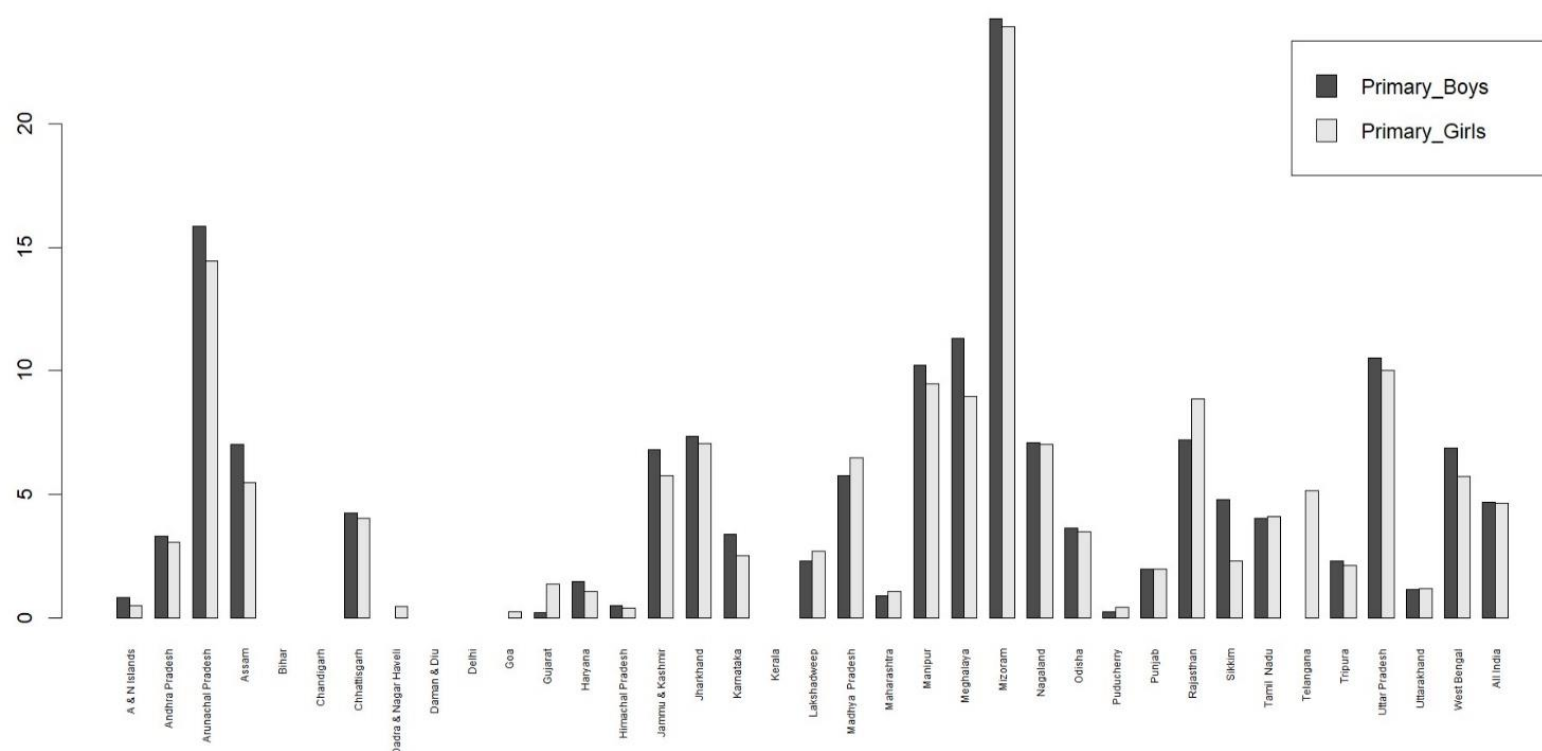
#Boys vs Girls 2012-13

```
compareable_data5 = rbind(year1213[,3],year1213[,4])
```

```
colnames(compareable_data5) = year1213[,1]
```

```
rownames(compareable_data5) = c("Primary_Boys", "Primary_Girls")
```

```
barplot(compareable_data5, beside=TRUE, legend=TRUE, cex.names = 0.5, las = 3)
```



Observations:

- All India: Boys Dropout rate is slightly greater than girls
- Gujrat, Lakshadweep, Madhya Pradesh, Maharashtra, Puducherry, Rajasthan, Tamil Nadu, Uttarakhand are states with Girls Dropout rate Higher than Boys Dropout Rate

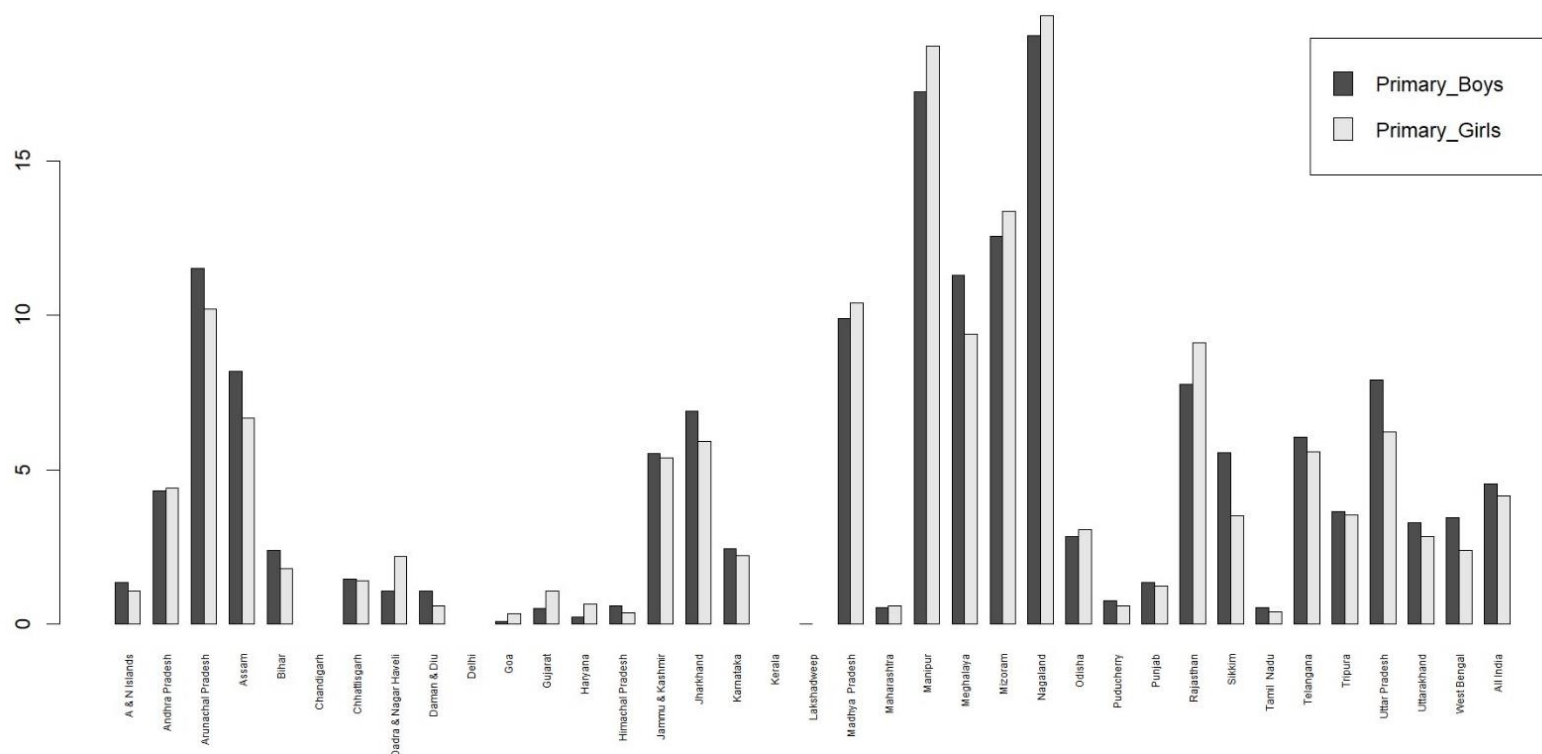
#Boys vs Girls 2013-14

```
compareable_data6 = rbind(year1314[,3],year1314[,4])
```

```
colnames(compareable_data6) = year1213[,1]
```

```
rownames(compareable_data6) = c("Primary_Boys", "Primary_Girls")
```

```
barplot(compareable_data6, beside=TRUE, legend=TRUE, cex.names = 0.5, las = 3)
```



Observations:

- All India: Boys Dropout rate is greater than girls
- Dadar & Nagar Haveli, Goa, Gujrat, Haryana, Madhya Pradesh, Maharashtra, Manipur, Mizoram, Nagaland, Odisha, Rajasthan are states with Girls Dropout rate Higher than Boys Dropout Rate

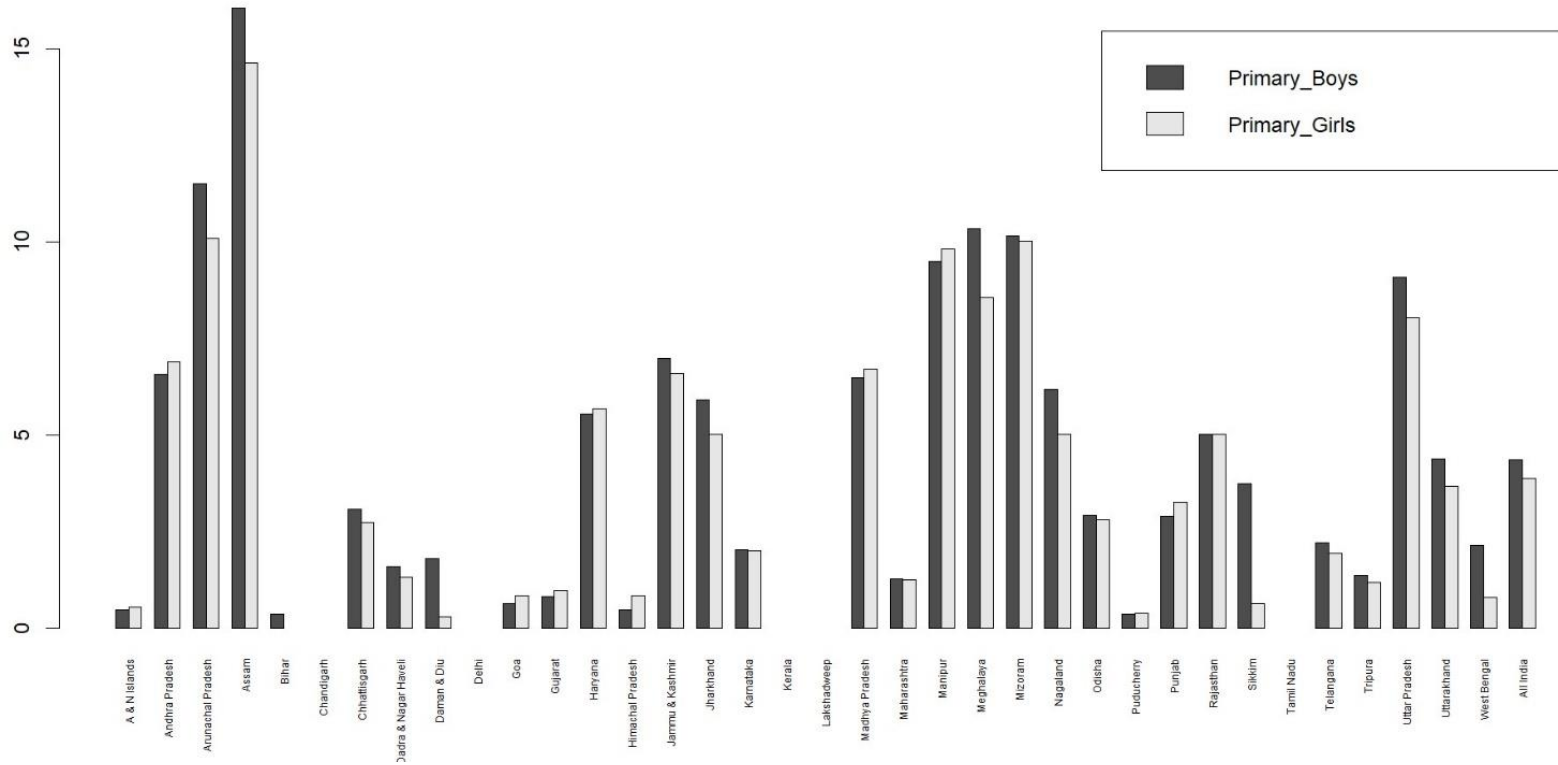
#Boys vs Girls 2014-15

```
compareable_data7 = rbind(year1415[,3],year1415[,4])
```

```
colnames(compareable_data7) = year1415[,1]
```

```
rownames(compareable_data7) = c("Primary_Boys", "Primary_Girls")
```

```
barplot(compareable_data7, beside=TRUE, legend=TRUE, cex.names = 0.5, las = 3)
```



Observations:

- All India: Boys Dropout rate is greater than girls
- A&N Islands, Goa, Gujrat, Haryana, Himachal Pradesh, Madhya Pradesh, Manipur, Puducherry, Punjab, Rajashtan are states with Girls Dropout rate Higher than Boys Dropout Rate

#### 4) Comparison b/w All India Boys and Girls Dropout Rate as they get promoted in classes

->For Primary School of Year 2012-13

#Choosing the columns which are to be plotted

```
boysIndex = c(3, 6, 9)
```

```
girlsIndex = boysIndex+1
```

#storing the data of choosed

```
boysclassData = year1213[37,boysIndex]
```

```
girlsclassData = year1213[37,girlsIndex]
```

#defining labels to be shown on both the axes

```
x_labels <- c("Primary", "Upper_Primary", "Secondary")
```

```
y_labels <- 0:15
```

#plotting the line graph with the 'girlclassdata' first

```
plot(1:3, girlsclassData, lwd=2, col="red", type="o", xaxt="n", yaxt="n", ylim=c(0,15), xlab="Class Groups", ylab="Students Dropped (%)", main="Graph to show the comparison of boys and girls dropout rates of different groups")
```

```
axis(side = 2, at= y_labels, labels = y_labels, las=2)
```

```
axis(side = 1, at=1:3, label = x_labels)
```

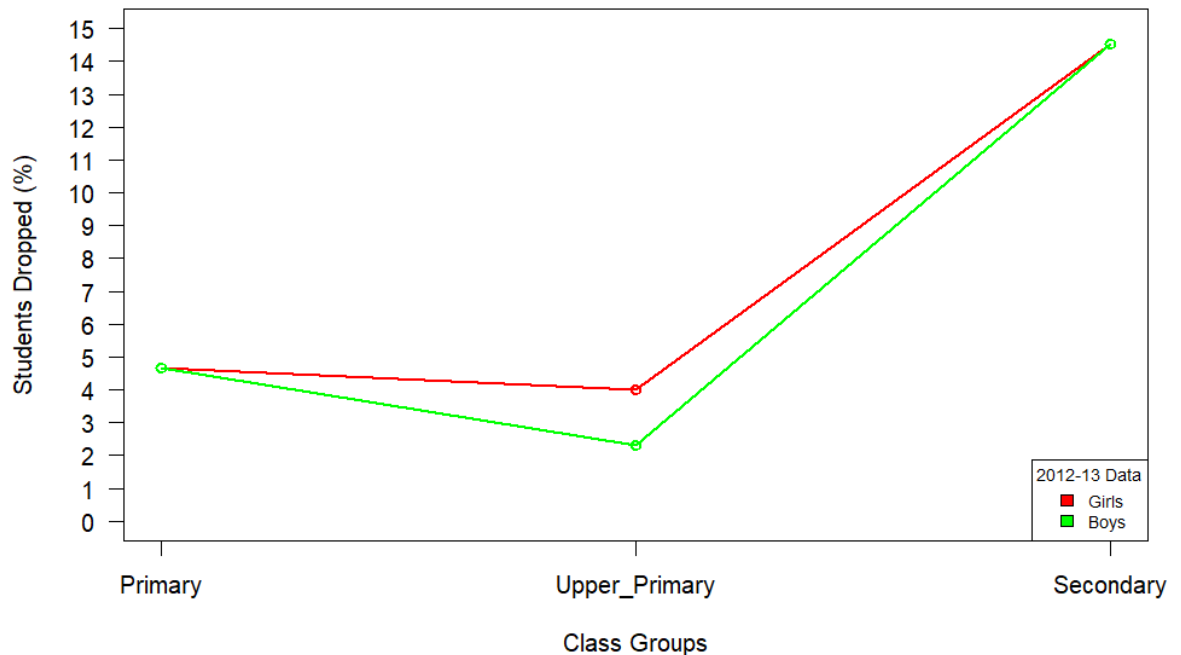
#plotting the second line of 'boysclassdata'

```
lines(seq(1,3), boysclassData, col="green", lwd="2", type="o" )
```

#defining the legends

```
legend("bottomright", title = "2012-13 Data", legend = c("Girls", "Boys"), fill = c("red", "green"), cex = 0.7)
```

**Graph to show the comparison of boys and girls dropout rates of different groups**



Observations:

- Drop rate Increases Significantly in Secondary School
- Girls have Significantly higher drop rate than Boys in Upper\_Primary School

**->For Primary School of Year 2013-14**

#Choosing the columns which are to be plotted

```
boysIndex = c(3, 6, 9, 12)
```

```
girlsIndex = boysIndex+1
```

#storing the data of choosed

```
boysclassData = year1314[37,boysIndex]
```

```
girlsclassData = year1314[37,girlsIndex]
```

#defining labels to be shown on both the axes

```
x_labels <- c("Primary", "Upper_Primary", "Secondary", "Higher_Secondary")
```

```
y_labels <- 0:18
```

#plotting the line graph with the 'girlclassdata' first

```
plot(1:4, girlsclassData, lwd=2, col="red", type="o", xaxt="n", yaxt="n", ylim=c(0,18), xlab = "Class Groups", ylab="Students Dropped (%)", main="Graph to show the comparison of boys and girls dropout rates of different groups")
```

```
axis(side = 2, at= y_labels, labels = y_labels, las=2)
```

```
axis(side = 1, at=1:4, label = x_labels)
```

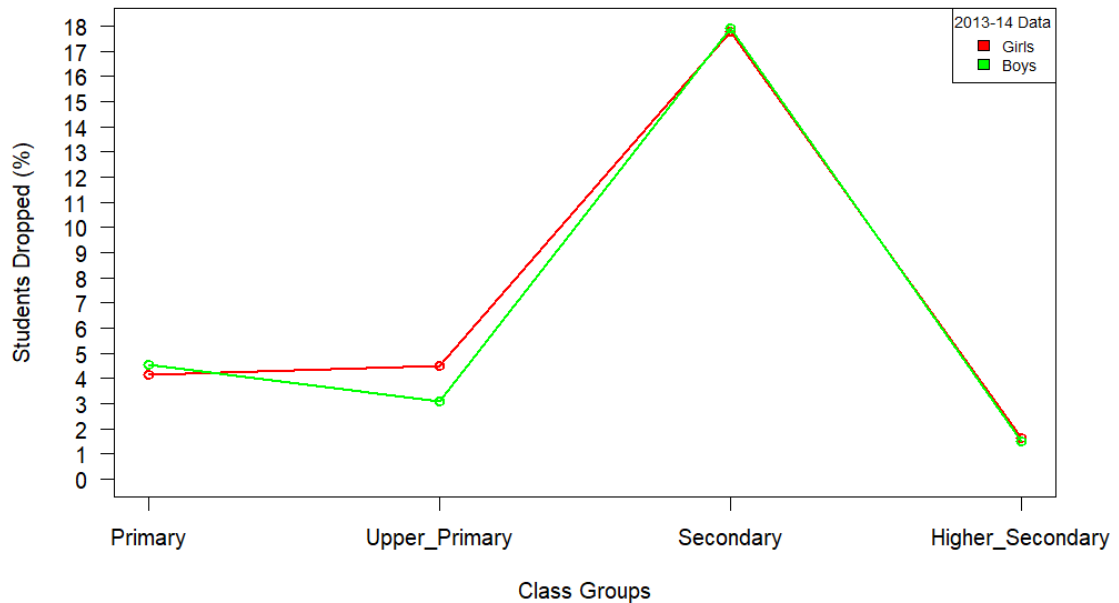
#plotting the second line of 'boysclassdata'

```
lines(seq(1,4), boysclassData, col="green", lwd="2", type="o" )
```

#defining the legends

```
legend("topright", title = "2013-14 Data", legend = c("Girls", "Boys"), fill = c("red", "green"), cex = 0.7)
```

**Graph to show the comparison of boys and girls dropout rates of different groups**



Observations:

- Drop rate Increases Significantly in Secondary School and Decreases Significantly in Higher\_Secondary School
- Girls have Significantly higher drop rate than Boys in Upper\_Primary School

**->For Primary School of Year 2014-15**

#Choosing the columns which are to be plotted

```
boysIndex = c(3, 6, 9)
```

```
girlsIndex = boysIndex+1
```

#storing the data of choosed

```
boysclassData = year1415[37,boysIndex]
```

```
girlsclassData = year1415[37,girlsIndex]
```

#defining labels to be shown on both the axes

```
x_labels <- c("Primary", "Upper_Primary", "Secondary")
```

```
y_labels <- 0:18
```

#plotting the line graph with the 'girlclassdata' first

```
plot(1:3, girlsclassData, lwd=2, col="red", type="o", xaxt="n", yaxt="n", ylim=c(0,18), xlab = "Class Groups", ylab="Students Dropped (%)", main="Graph to show the comparison of boys and girls dropout rates of different groups")
```

```
axis(side = 2, at= y_labels, labels = y_labels, las=2)
```

```
axis(side = 1, at=1:3, label = x_labels)
```

#plotting the second line of 'boysclassdata'

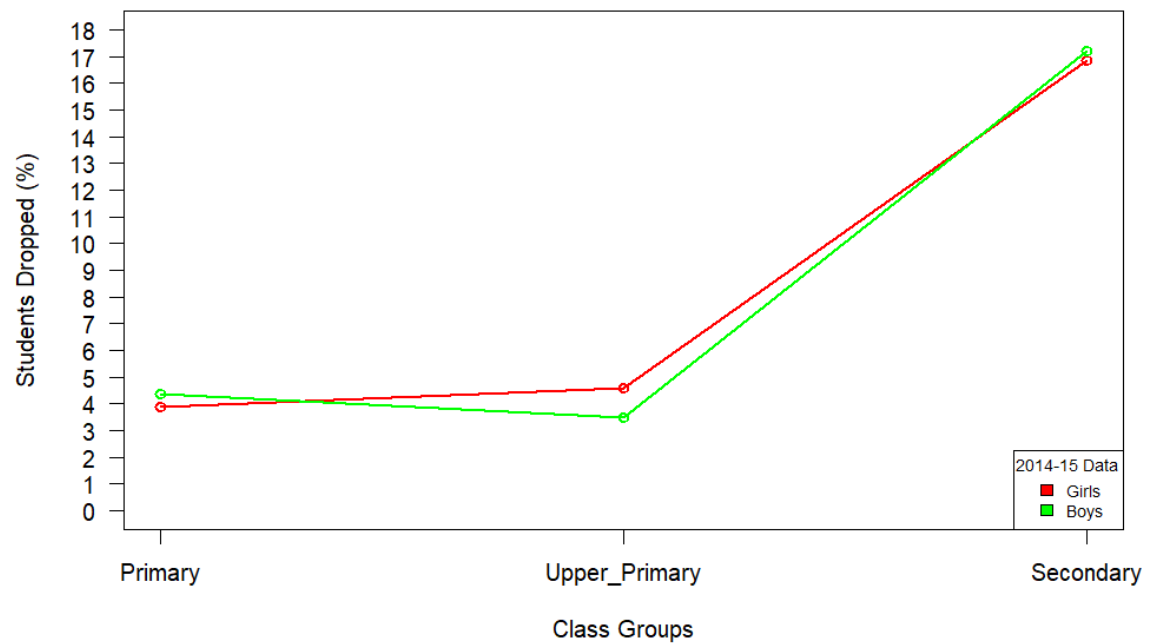
```
lines(seq(1,3), boysclassData, col="green", lwd="2", type="o" )
```

#defining the legends

```
legend("bottomright", title = "2014-15 Data", legend = c("Girls", "Boys"), fill = c("red", "green"), cex = 0.7)
```



**Graph to show the comparison of boys and girls dropout rates of different groups**



**Observations:**

- Drop rate Increases Significantly in Secondary School and Decreases Significantly in Higher\_Secondary School
- Girls have Significantly higher drop rate than Boys in Upper\_Primary School, while the difference has decreased compared to previous years

## 5) Normality Check for Primary School Dropout rate

For Normality Check we will use Histogram and density plot

->For Primary School of Year 2012-13

#Creating a subset for the Primary\_Total Data for year 2012-13

```
Primary_year1213<- subset(year1213,select = c(State_UT, Primary_Total))
```

#removing the row of All India data

```
Primary_year1213 <- Primary_year1213[-37, ]
```

#removing the NA(s)

```
Primary_year1213 <- na.omit(Primary_year1213)
```

#storing the values of Primary\_Total data in variable 'values'

```
values = Primary_year1213[,2]
```

#Plotting a histogram using the data in variable 'values'

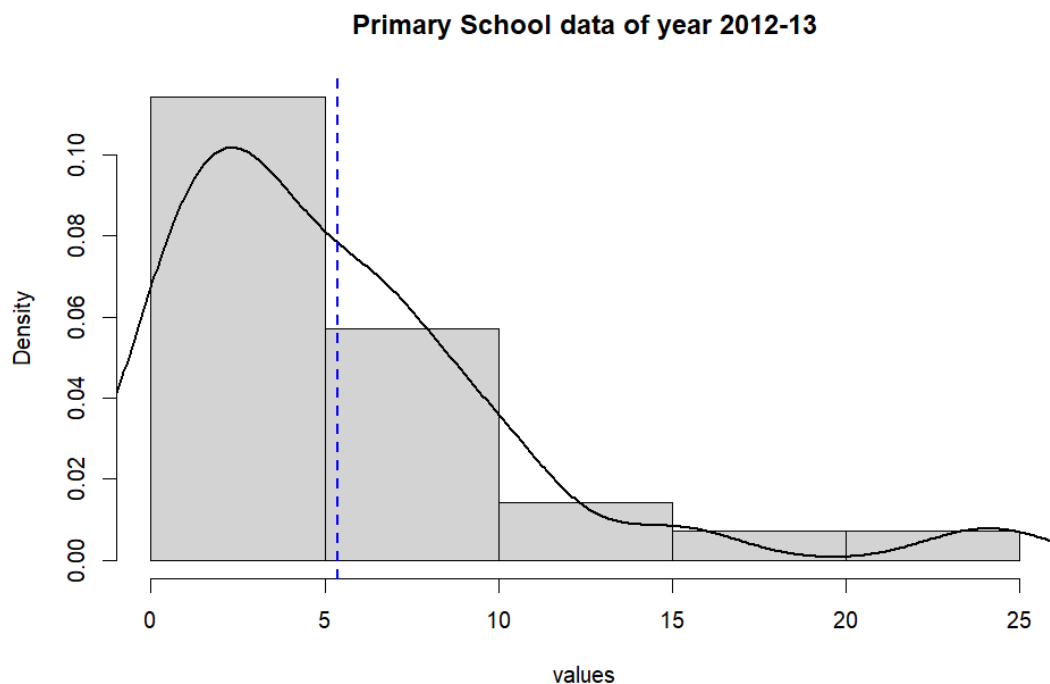
```
hist(values, freq = FALSE, main="Primary School data of year 2012-13")
```

#creating a curve using density function on 'values'

```
lines(density(values),lwd=2)
```

#Plotting the mean as a line parallel to y axis

```
abline(v = mean(values), col = 'blue', lwd = 2, lty = 2)
```



Observations:

- Mizoram is only state with dropout rate > 20%
- Since this curve is not Bell Shaped Symmetrically distributed it is not Normal Distributed

**->For Primary School of Year 2013-14**

#Creating a subset for the Primary\_Total Data for year 2013-14

```
Primary_year1314<- subset(year1314,select = c(State_UT, Primary_Total))
```

#removing the row of All India data

```
Primary_year1314 <- Primary_year1314[-37, ]
```

#removing the NA(s)

```
Primary_year1314 <- na.omit(Primary_year1314)
```

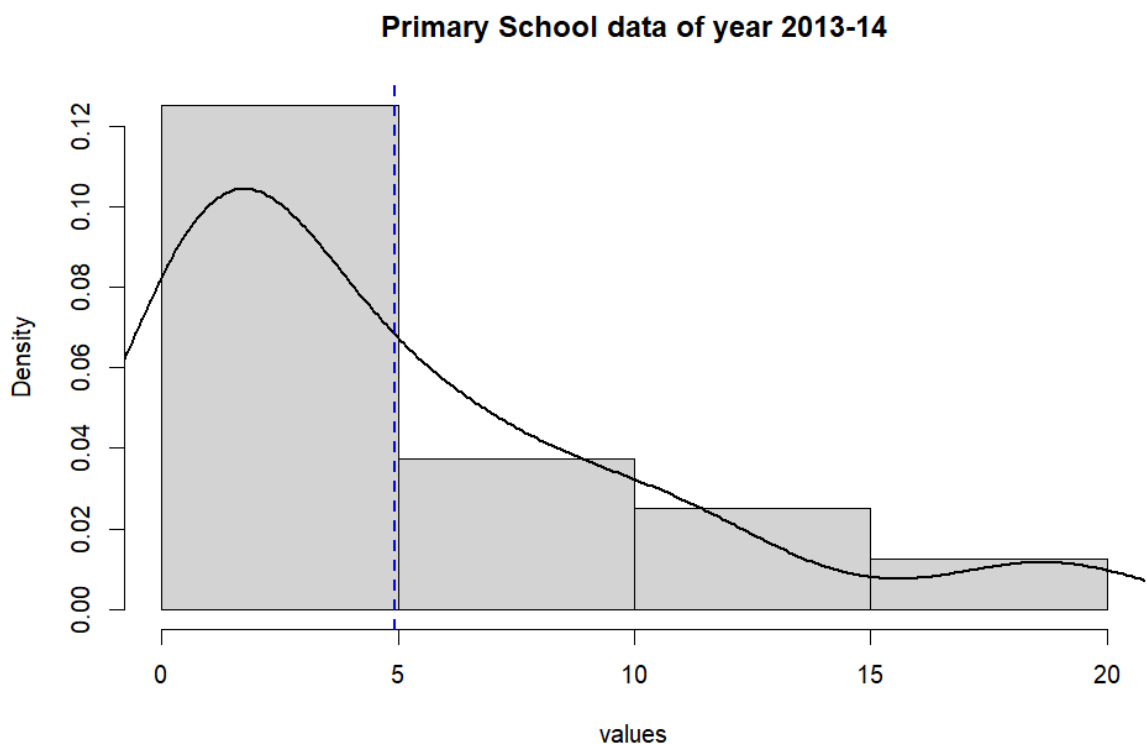
#storing the values of Primary\_Total data in variable 'values'

```
values = Primary_year1314[,2]
```

```
hist(values, freq = FALSE, main="Primary School data of year 2013-14")
```

```
lines(density(values),lwd=2)
```

```
abline(v = mean(values), col = 'blue', lwd = 2, lty = 2)
```



Observations:

- Since this curve is not Bell Shaped Symmetrically distributed it is not Normal Distributed
- Nagaland and Manipur are the only states with dropout rate >15%

#### ->For Primary School of Year 2014-15

#Creating a subset for the Primary\_Total Data for year 2014-15

```
Primary_year1415<- subset(year1415,select = c(State_UT, Primary_Total))
```

#removing the row of All India data

```
Primary_year1415 <- Primary_year1415[-37, ]
```

#removing the NA(s)

```
Primary_year1415 <- na.omit(Primary_year1415)
```

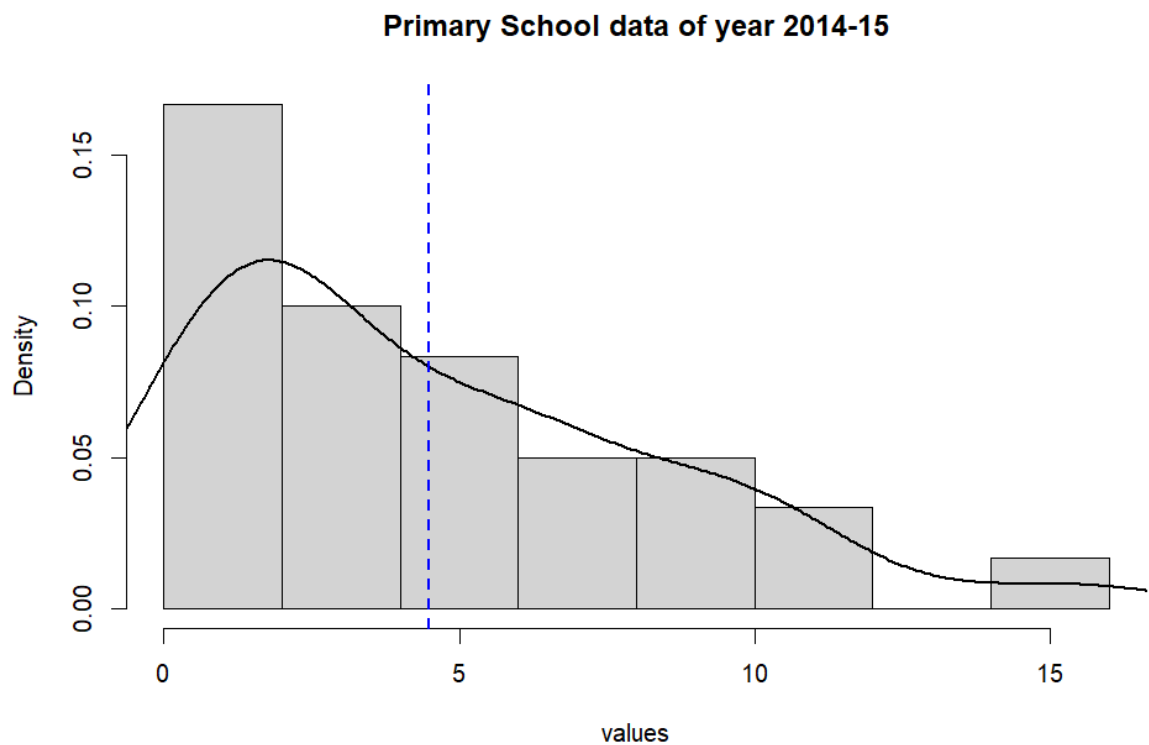
#storing the values of Primary\_Total data in variable 'values'

```
values = Primary_year1415[,2]
```

```
hist(values, freq = FALSE, main="Primary School data of year 2014-15")
```

```
lines(density(values),lwd=2)
```

```
abline(v = mean(values), col = 'blue', lwd = 2, lty = 2)
```



Observations:

- Since this curve is not Bell Shaped Symmetrically distributed it is not Normal Distributed
- Assam has the highest dropout rate >15%
- For all 3 years Mean is around 5 and decreasing every year