CSE-3020

Data Visualization (J component)

~ Fall Semester 2021-22 ~

SUBMITTED TO:

LYDIA JANE G

SLOT – L43+L44

SUBMITTED BY:

Arush Saxena (20BCE2106)

Review 1:-

Dataset Details:-

a. URL:-

https://github.com/Arush413/Analysis-of-Primary-Indian-Census-2011/blob/main/Primary_Census_Abstract_Total_Table_For_India.csv

b. Reading the data:-

```
r=read.csv("W:/Study Material/DV/Project/Primary_Census_Abstract_Total_Table_For_India.csv")
```

For getting number of rows and columns:-

```
nrow(r)
```

Output:-

108

```
ncol(r)
```

Output:-

94

c. Attributes Definition (Attribute,Definition):-

```
State,State Code

District,District Code

Subdistt,Subdistrict code

Town/Village,Town/Village Code

EB,Enumeration Block

Level,Level
```

```
Name,Name

TRU,Total/Rural/Urban

No_HH,No of Household

TOT_P,Total population Person

TOT_M,Total population Male

TOT_F,Total population Female

P_06,Population in the age group 0-6 Person

M_06,Population in the age group 0-6 Male

F_06,Population in the age group 0-6 Female

P_SC,Scheduled Castes population Person

M_SC,Scheduled Castes population Male

F_SC,Scheduled Castes population Female

P_ST,Scheduled Tribes population Person

M_ST,Scheduled Tribes population Male

F_ST,Scheduled Tribes population Female

P_LIT,Literates population Person

M_LIT,Literates population Male

F_LIT,Literates population Female

P_ILL,Illiterate Persons
```

M_ILL,Illiterate Male

F_ILL,Illiterate Female

TOT_WORK_P,Total Worker Population Person

TOT_WORK_M,Total Worker Population Male

TOT_WORK_F,Total Worker Population Female

MAINWORK_P,Main Working Population Person

MAINWORK_M,Main Working Population Male

MAINWORK_F,Main Working Population Female

MAIN_CL_P,Main Cultivator Population Person

MAIN_CL_M,Main Cultivator Population Male

MAIN_CL_F,Main Cultivator Population Female

MAIN_AL_P,Main Agricultural Labourers Population Person

MAIN_AL_M,Main Agricultural Labourers Population Male

MAIN_AL_F,Main Agricultural Labourers Population Female

MAIN_HH_P,Main Household Industries Population Person

MAIN_HH_M,Main Household Industries Population Male

MAIN_HH_F,Main Household Industries Population Female

MAIN_OT_P,Main Other Workers Population Person

MAIN_OT_M,Main Other Workers Population Male

MAIN_OT_F,Main Other Workers Population Female

MARGWORK_P,Marginal Worker Population Person

MARGWORK_M,Marginal Worker Population Male

MARGWORK_F,Marginal Worker Population Female

MARG_CL_P,Marginal Cultivator Population Person

MARG_CL_M,Marginal Cultivator Population Male

MARG_CL_F,Marginal Cultivator Population Female

MARG_AL_P,Marginal Agriculture Labourers Population Person

MARG_AL_M,Marginal Agriculture Labourers Population Male

MARG_AL_F,Marginal Agriculture Labourers Population Female

MARG_HH_P,Marginal Household Industries Population Person

MARG_HH_M,Marginal Household Industries Population Male

MARG_HH_F,Marginal Household Industries Population Female

MARG_OT_P,Marginal Other Workers Population Person

MARG_OT_M,Marginal Other Workers Population Male

MARG_OT_F,Marginal Other Workers Population Female

MARGWORK_3_6_P,Marginal Worker Population 3-6 Person

MARGWORK_3_6_M,Marginal Worker Population 3-6 Male

MARGWORK_3_6_F,Marginal Worker Population 3-6 Female

MARG_CL_3_6_P,Marginal Cultivator Population 3-6 Person

MARG_CL_3_6_M,Marginal Cultivator Population 3-6 Male

MARG_CL_3_6_F,Marginal Cultivator Population 3-6 Female

MARG_AL_3_6_P,Marginal Agriculture Labourers Population 3-6 Person

MARG_AL_3_6_M,Marginal Agriculture Labourers Population 3-6 Male

MARG_AL_3_6_F,Marginal Agriculture Labourers Population 3-6 Female

MARG_HH_3_6_P,Marginal Household Industries Population 3-6 Person

MARG_HH_3_6_M,Marginal Household Industries Population 3-6 Male

MARG_HH_3_6_F,Marginal Household Industries Population 3-6 Female

MARG_OT_3_6_P,Marginal Other Workers Population Person 3-6 Person

MARG_OT_3_6_M,Marginal Other Workers Population Person 3-6 Male

MARG_OT_3_6_F,Marginal Other Workers Population Person 3-6 Female

MARGWORK_0_3_P,Marginal Worker Population 0-3 Person

MARGWORK_0_3_M,Marginal Worker Population 0-3 Male

MARGWORK_0_3_F,Marginal Worker Population 0-3 Female

MARG_CL_0_3_P,Marginal Cultivator Population 0-3 Person

MARG_CL_0_3_M,Marginal Cultivator Population 0-3 Male

MARG_CL_0_3_F,Marginal Cultivator Population 0-3 Female

MARG_AL_0_3_P,Marginal Agriculture Labourers Population 0-3 Person

MARG_AL_0_3_M,Marginal Agriculture Labourers Population 0-3 Male

MARG_AL_0_3_F,Marginal Agriculture Labourers Population 0-3 Female

MARG_HH_0_3_P,Marginal Household Industries Population 0-3 Person

MARG_HH_0_3_M,Marginal Household Industries Population 0-3 Male

MARG_HH_0_3_F,Marginal Household Industries Population 0-3 Female

MARG_OT_0_3_P,Marginal Other Workers Population 0-3 Person

MARG_OT_0_3_M,Marginal Other Workers Population 0-3 Male

MARG_OT_0_3_F,Marginal Other Workers Population 0-3 Female

NON_WORK_P,Non Working Population Person

NON_WORK_M,Non Working Population Male

NON_WORK_F,Non Working Population Female

Data Abstraction:-

The selected dataset describes about Indian Census of the year 2011 which can also be said as a comprehensive topic based on policies made by the central government of India. The Indian population stood at 1.21 billion which is around 17.7% more than that of the 2001 census.

Now, talking about the data semantics, we have divided the data into:-

Row-wise (Divided as per area):-

We divided the data as per the geographical areas initiating from the country level and then moving down to identifying the areas as either rural, urban or both.

Firstly, we have the level attribute which divides data into either from the whole country or from a state.

Then, data is categorized as per each state and further it is divided into urban, rural or total population.

Hence, the numerical data we have, is divided geographically which can also be plotted as a map/network.

Column-wise (Divided as per attribute):-

The selected data is a numerical data type which comes under discrete category which can't be precisely measured but can be counted as per the figures provided in each column of the table.

The column-wise division is done as per the different aspects of the Indian population, i.e., Literacy rate, Health care, working population, Households, caste, etc.

It is further distributed according to three parameters which are 'total', 'male' and 'female' counts of each aspect.

This dataset of census can actually be read in the form of either table as discussed above (column and row analysis) or a Data Structure namely tree where the root node is our country, India and its child nodes are the states which have further children as the type of areas which are urban and rural and both.

Hence, we can conclude that the data we have taken for analysis is actually divided into Indian geographical areas where attributes have been defined as per the Indian aspects which are further categorically divided into total counts and male & female counts for each attribute division.

Task Abstraction:-

1.What is the population per state of our country according to the 2011 census data? Which are the most and least populous states?

Solution:-

The total population of each state has to be found out from the data file which is done by using the code:-

```
pop_state=subset(r,Level=="STATE"&TRU=="Total",select=c(TOT_P,Name))

pop_state
```

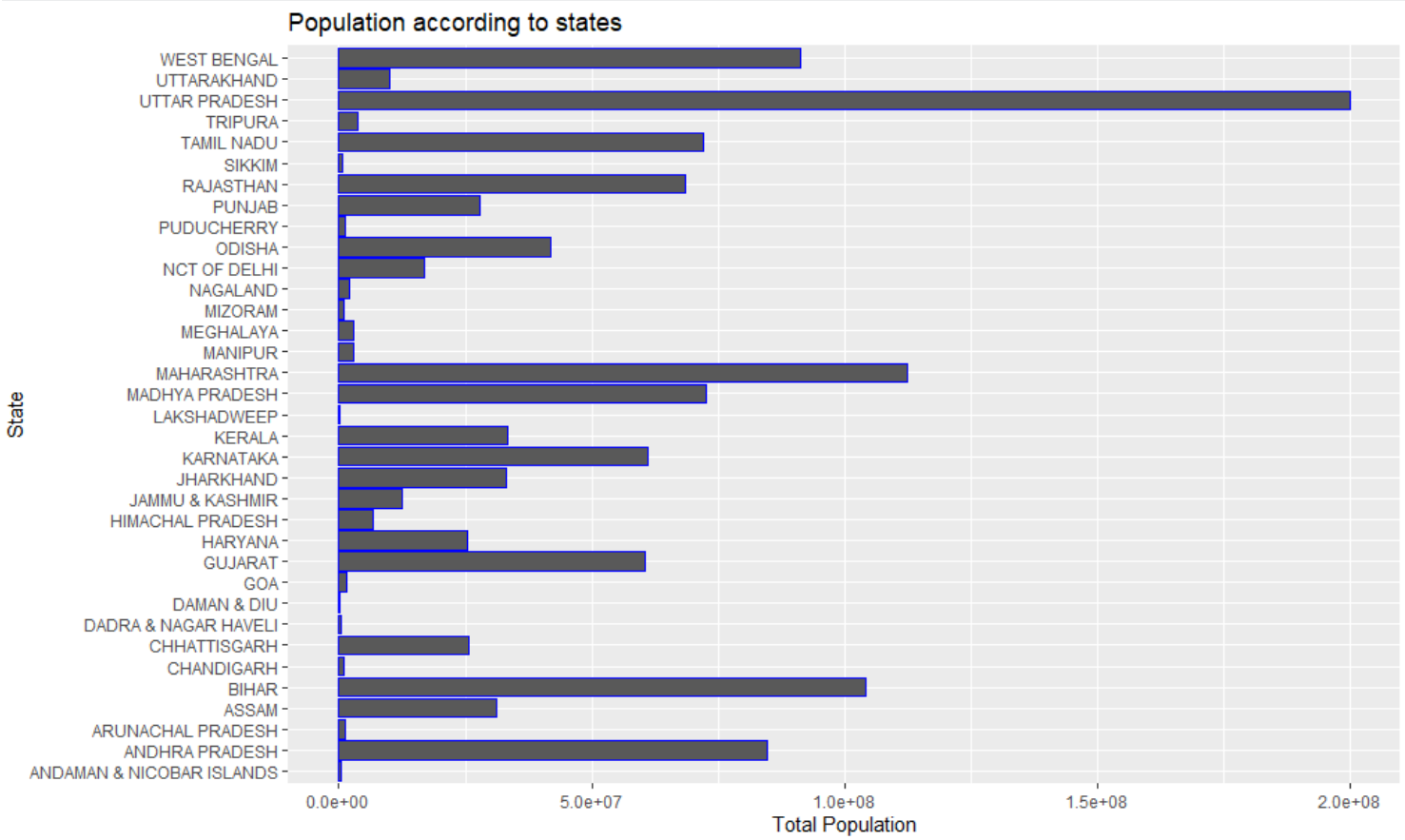
Output:-

	TOT_P	Name
4	12541302	JAMMU & KASHMIR
7	6864602	HIMACHAL PRADESH
10	27743338	PUNJAB
13	1055450	CHANDIGARH
16	10086292	UTTARAKHAND
19	25351462	HARYANA
22	16787941	NCT OF DELHI
25	68548437	RAJASTHAN
28	199812341	UTTAR PRADESH
31	104099452	BIHAR
34	610577	SIKKIM
37	1383727	ARUNACHAL PRADESH
40	1978502	NAGALAND
43	2855794	MANIPUR
46	1097206	MIZORAM
49	3673917	TRIPURA
52	2966889	MEGHALAYA
55	31205576	ASSAM
58	91276115	WEST BENGAL
61	32988134	JHARKHAND
64	41974218	ODISHA
67	25545198	CHHATTISGARH
70	72626809	MADHYA PRADESH
73	60439692	GUJARAT
76	243247	DAMAN & DIU
79	343709	DADRA & NAGAR HAVELI
82	112374333	MAHARASHTRA
85	84580777	ANDHRA PRADESH
88	61095297	KARNATAKA
91	1458545	GOA
94	64473	LAKSHADWEEP
97	33406061	KERALA
100	72147030	TAMIL NADU
103	1247953	PUDUCHERRY
106	380581	ANDAMAN & NICOBAR ISLANDS

A barplot of the new “state vs population” subset can be made by the code:-

```
pop_state_graph=ggplot(pop_state,aes(x=TOT_P,y=Name))+geom_bar(stat="identity",colour="blue")+xlab("Total Population")+ylab("State")+ggtitle("Population according to states")

pop_state_graph
```

Output:-



From the data and the graph, it is evident that the most populous state is Uttar Pradesh and least populous state is Lakshadweep.

2.What is the status of sex ratio according to the 2011 census data?

Solution:-

The total male and female population has to be found out for each state which can be done using the code:-

```
sex_ratio=subset(r,Level=="STATE"&TRU=="Total",select=c(Name,TOT_M,TOT_F))

sex_ratio
```

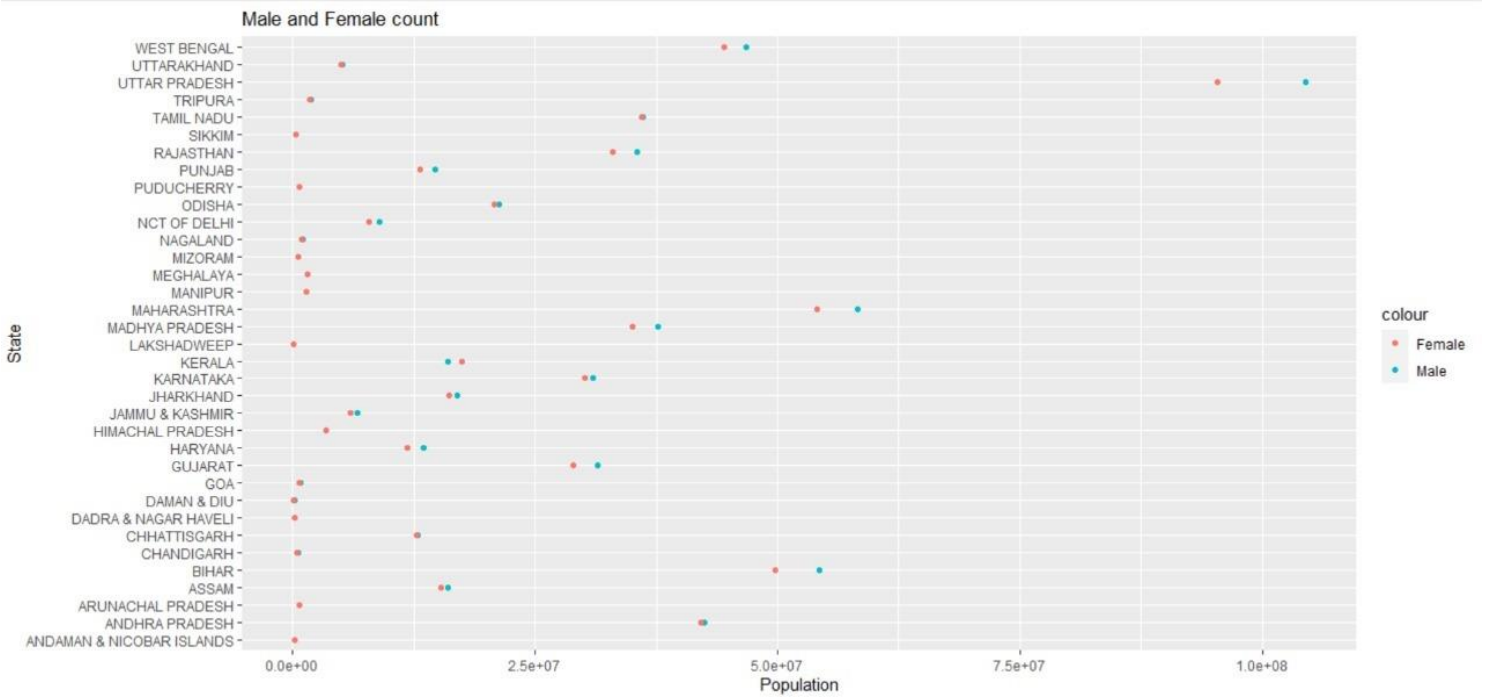
Output:-

	Name	TOT_M	TOT_F
4	JAMMU & KASHMIR	6640662	5900640
7	HIMACHAL PRADESH	3481873	3382729
10	PUNJAB	14639465	13103873
13	CHANDIGARH	580663	474787
16	UTTARAKHAND	5137773	4948519
19	HARYANA	13494734	11856728
22	NCT OF DELHI	8987326	7800615
25	RAJASTHAN	35550997	32997440
28	UTTAR PRADESH	104480510	95331831
31	BIHAR	54278157	49821295
34	SIKKIM	323070	287507
37	ARUNACHAL PRADESH	713912	669815
40	NAGALAND	1024649	953853
43	MANIPUR	1438586	1417208
46	MIZORAM	555339	541867
49	TRIPURA	1874376	1799541
52	MEGHALAYA	1491832	1475057
55	ASSAM	15939443	15266133
58	WEST BENGAL	46809027	44467088
61	JHARKHAND	16930315	16057819
64	ODISHA	21212136	20762082
67	CHHATTISGARH	12832895	12712303
70	MADHYA PRADESH	37612306	35014503
73	GUJARAT	31491260	28948432
76	DAMAN & DIU	150301	92946
79	DADRA & NAGAR HAVELI	193760	149949
82	MAHARASHTRA	58243056	54131277
85	ANDHRA PRADESH	42442146	42138631
88	KARNATAKA	30966657	30128640
91	GOA	739140	719405
94	LAKSHADWEEP	33123	31350
97	KERALA	16027412	17378649
100	TAMIL NADU	36137975	36009055
103	PUDUCHERRY	612511	635442
106	ANDAMAN & NICOBAR ISLANDS	202871	177710

A scatterplot can be made of the newly extracted data by using the code:-

```
male_female_plot=ggplot(sex_ratio)+geom_point(aes(TOT_M,Name,col='Male'))
+geom_point(aes(TOT_F,Name,col='Female'))+xlab('Population')+ylab('State')+ggt
itle('Male and Female count')

male_female_plot
```

Output:-



It is clear from the graph that the sex ratio in most states of India is very low (depends on the distance between the gender plot of each state).

3.What is the share of main working population between the two genders in different types of regions of India according to 2011 census data?

Solution:-

We have to obtain the male and female working population in rural and urban areas of India which can be done using the code:-

```
tru_work=subset(r,Level=="India",select=c(TRU,TOT_WORK_M,TOT_WORK_F))

tru_work
```

Output:-

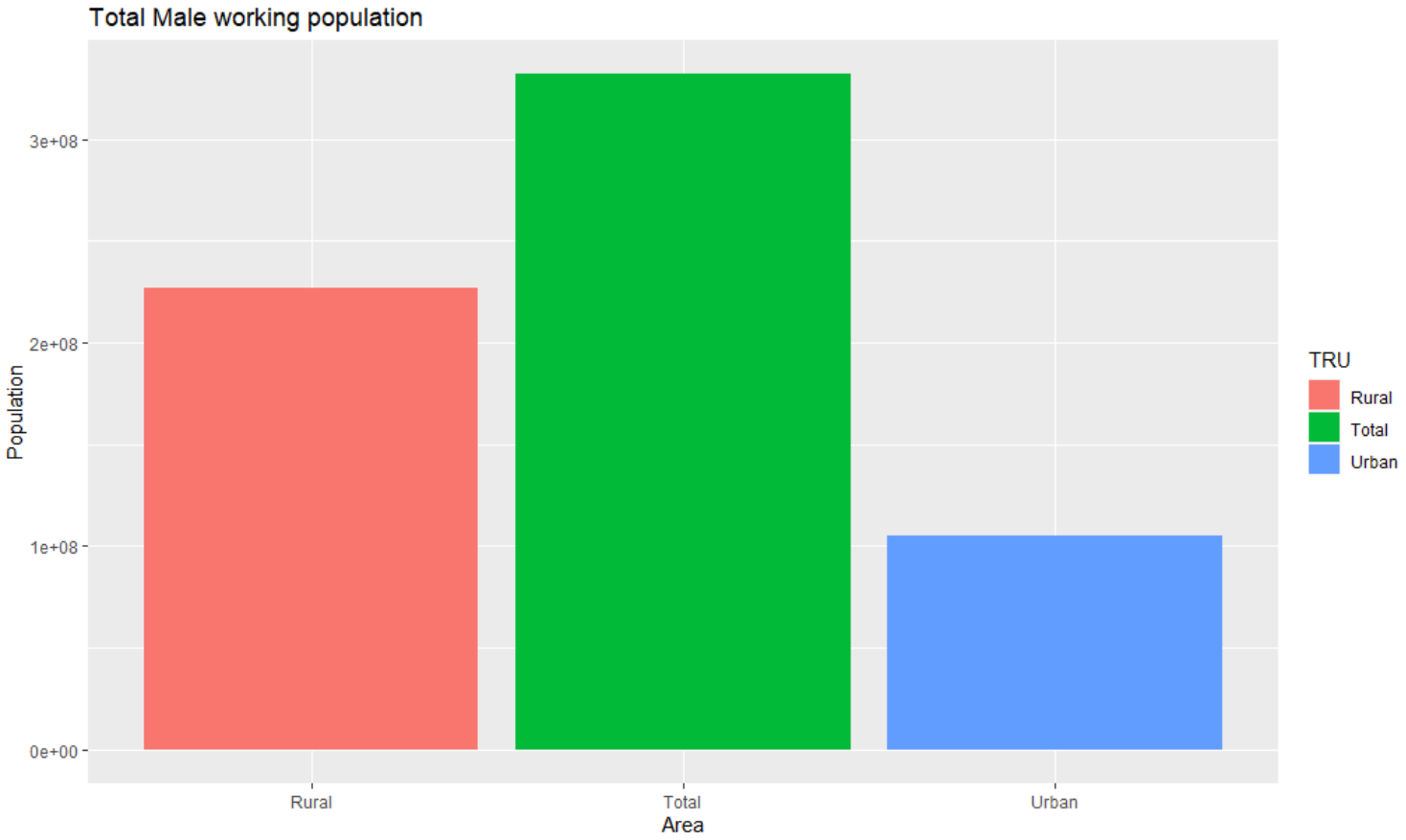
```
     TRU TOT_WORK_M TOT_WORK_F
1  Total  331939875  149948993
2  Rural  226837013  121906079
3  Urban  105102862   28042914
```

We can plot the barplot for each gender working in rural and urban areas using the codes:-

For Male:-

```
tru_work_graph=ggplot(tru_work, aes(TRU,TOT_WORK_M,
fill=TRU))+geom_bar(stat="identity")+xlab("Area")+ylab("Population")+ggtitle("To
tal Male working population")

tru_work_graph
```

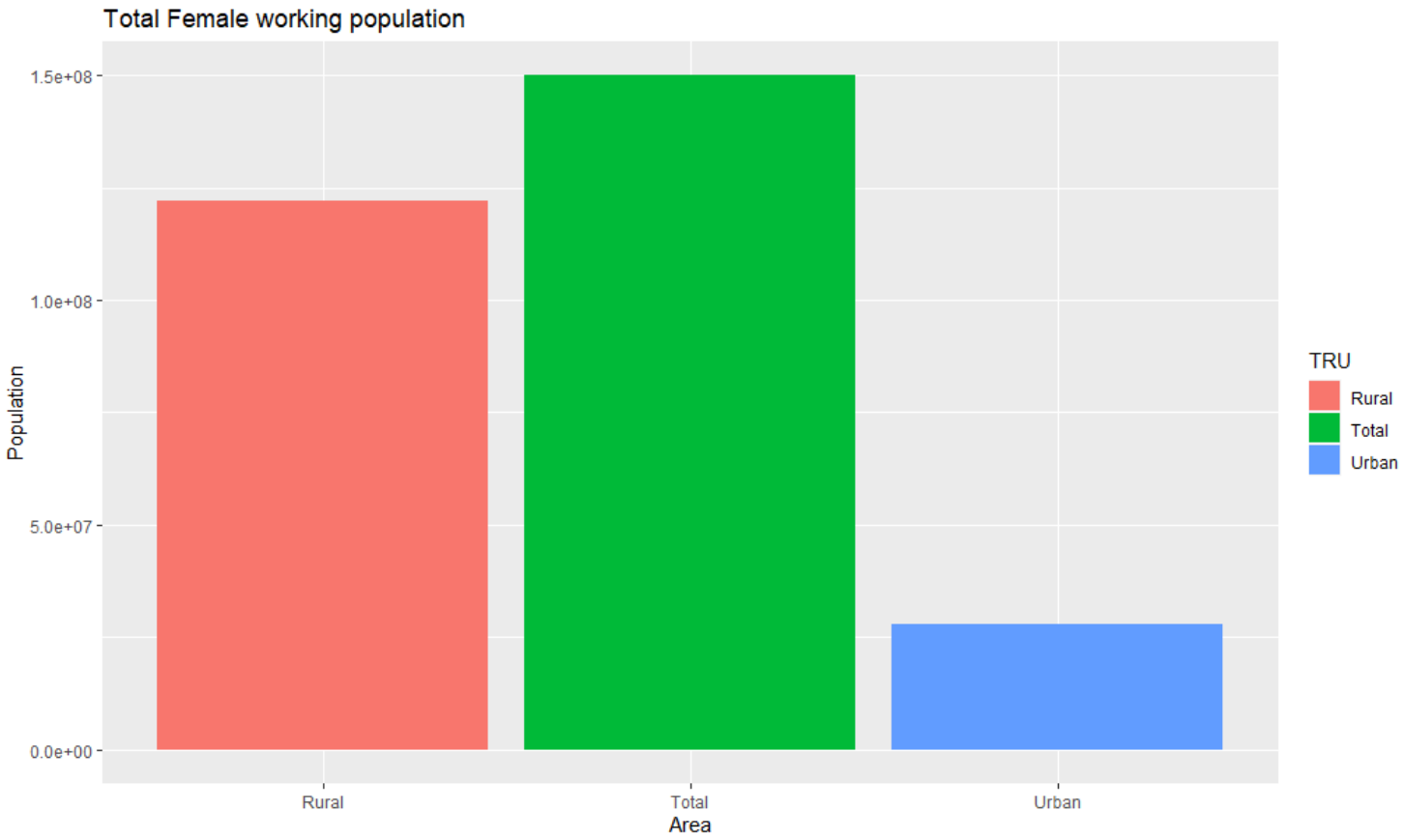
Output:-



In a similar way, we do the same for Female working population:-

```
tru_work_graph1=ggplot(tru_work, aes(TRU,TOT_WORK_F, fill=TRU))+geom_bar(stat="identity")+xlab("Area")+ylab("Population")+ggtitle("Total Female working population")

tru_work_graph1
```

Output:-



It can be seen from the graph that the rural working population is far greater than urban working population irrespective of the gender.

4.How is the main Indian working population divided into different categories of occupations according to the 2011 Indian census data?

Solution:-

We have to obtain the main working population of the different types of occupation sectors like cultivation, agricultural, household and other working population. It can be achieved using the following code:-

```
main_work_share=
subset(r,Level=="India"&TRU=="Total",select=c(MAIN_CL_P,MAIN_AL_P,MAIN_HH_P,MAIN_OT_P))

main_work_share
```

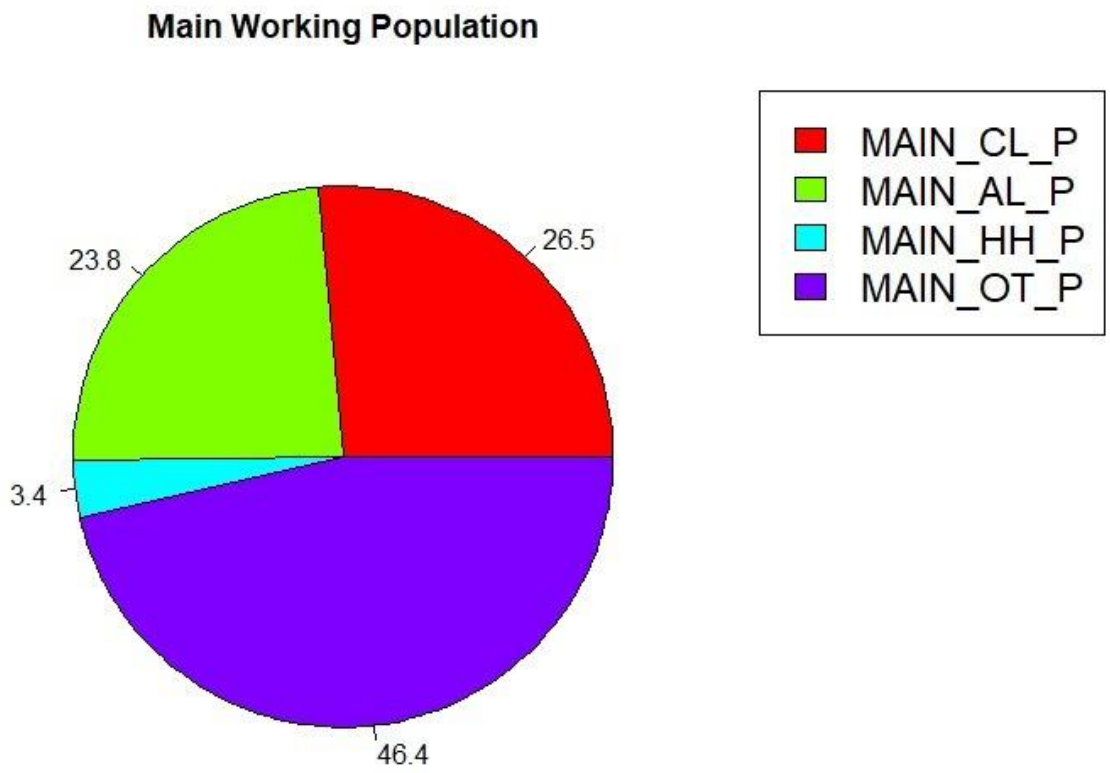
Output:-

```
  MAIN_CL_P MAIN_AL_P MAIN_HH_P MAIN_OT_P
1  95942413  86168706  12332802 168121650
```

The percentages of each sector along with their pie chart can be obtained using the codes:-

```
work_share_plot_data=c(main_work_share$MAIN_CL_P,main_work_share$MAIN_AL_P,main_work_share$MAIN_HH_P,main_work_share$MAIN_OT_P)

piepercent=round(100*work_share_plot_data/sum(work_share_plot_data),1)

work_share_plot=pie(work_share_plot_data,labels=piepercent,main=("Main Working Population"),col=rainbow(length(work_share_plot_data)))

legend("topright",c("MAIN_CL_P","MAIN_AL_P","MAIN_HH_P","MAIN_OT_P"),cex=1.5,fill=rainbow(length(work_share_plot_data)))
```

Output:-



Hence, the share of main working population in all sectors is obtained and major population is involved in OT (other working population).

5. What is the share of non working population between the two genders in different types of regions of India according to 2011 census data?

Solution:-

We have to get the total, male and female non-working population in urban and rural parts of India which can be done using the code:-

```
nonwork=subset(r,Level=="India",select=c(TRU,NON_WORK_P,NON_WORK_M,NON_WORK_F))

nonwork
```

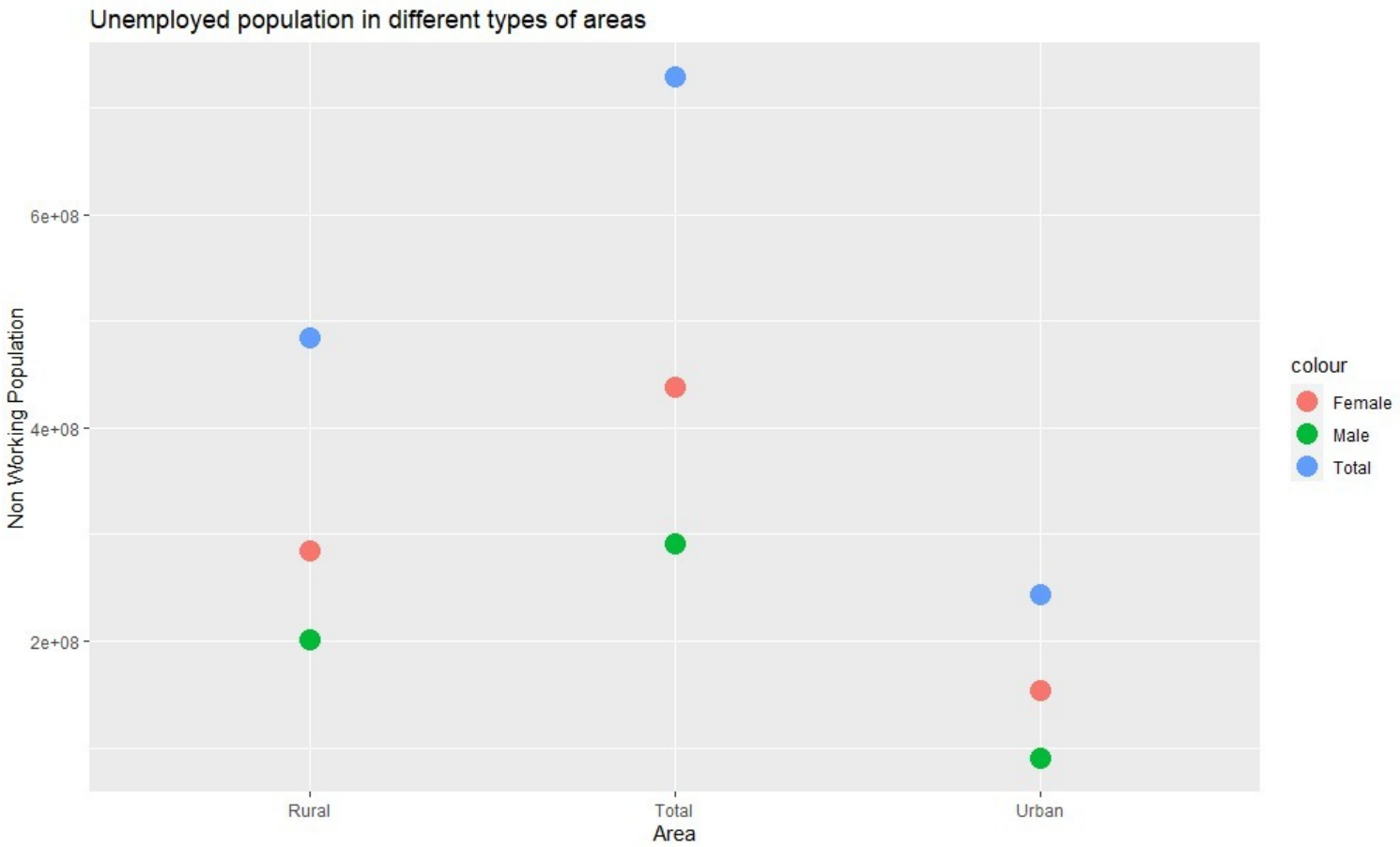
Output:-

```
     TRU NON_WORK_P NON_WORK_M NON_WORK_F
1  Total  728966109  291330383  437635726
2  Rural  485005760  200944045  284061715
3  Urban  243960349   90386338  153574011
```

The stripchart for the same can be obtained using:-

```
nonwork_plot=ggplot(nonwork)+geom_point(size=5,aes(TRU,NON_WORK_P,col="Total"))+geom_point(size=5,aes(TRU,NON_WORK_M,col="Male"))+geom_point(size=5,aes(TRU,NON_WORK_F,col="Female"))+xlab("Area")+ylab("Non Working Population")+ggtitle("Unemployed population in different types of areas")

nonwork_plot
```

Output:-



It is evident from the plot that unemployment in rural areas is far more than that in urban areas. Also, the unemployment for female population is more than male irrespective of the area.

6.How is the total Indian population divided between rural and urban regions of India according to 2011 Indian census data?

Solution:-

The total population can be segregated in terms of urban and rural India using the following code:-

```
tru = subset(r,Level=="India",select=c(TOT_P,TRU))

tru
```

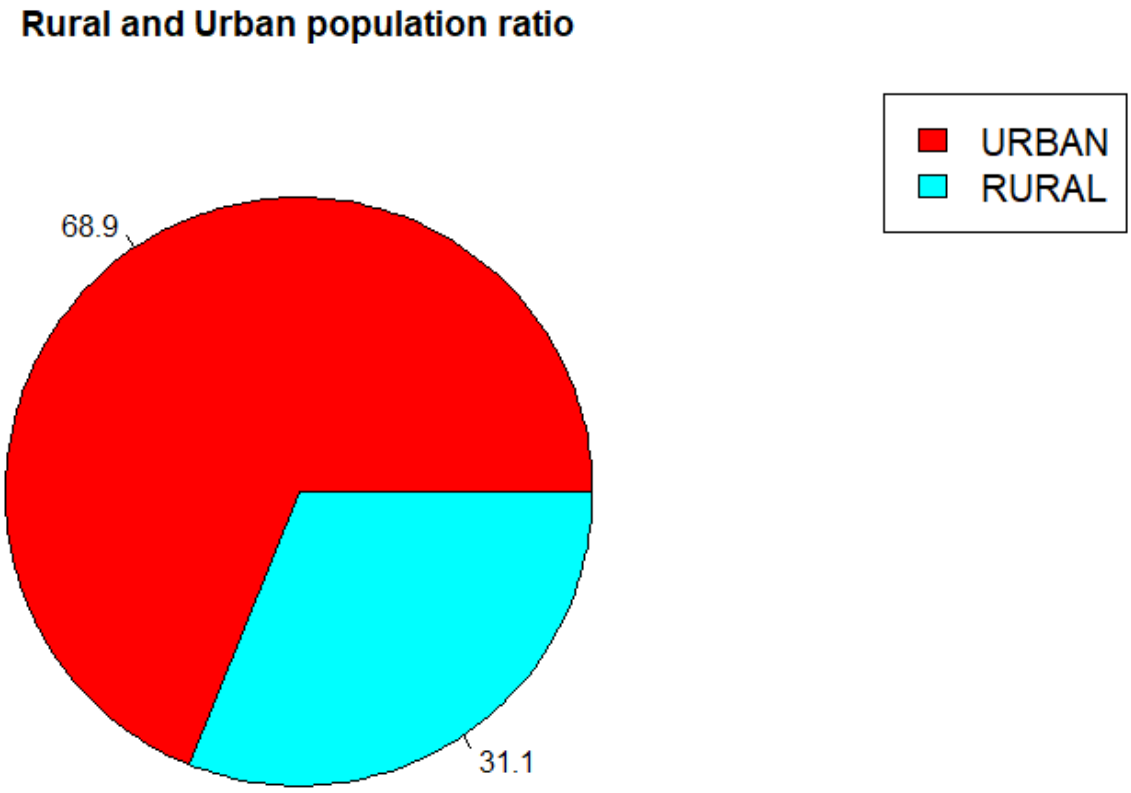
Output:-

```
      TOT_P   TRU
1 1210854977 Total
2  833748852 Rural
3  377106125 Urban
```

The percentage of this distribution can be visualized using pie chart which can be obtained using the codes:-

```
r11 <-  c(833748852,377106125)

labelsr11 <-  c("Rural", "Urban")

piepercent<- round(100 * r11 / 1210854977, 1)

pie(r11, labels = piepercent, main = "FRACTION OF URBAN AND RURAL
POPULATION", col = rainbow(length(r11)))

legend("topright", c("URBAN", "RURAL"),cex = 1.3, fill = rainbow(length(r11)))
```

Output:-



Hence, approximately 68.9% population of India resides in urban areas.

7.What is the literate population in each state of India according to the last census? Which are the most and least literate states in India?

Solution:-

Here, the total literate population of all the states has to be found out from the data. It can be done using the code:-

literate_state=subset(r,Level=="STATE"&TRU=="Total",select=c(Name,P_LIT))

literate_state

Output:-

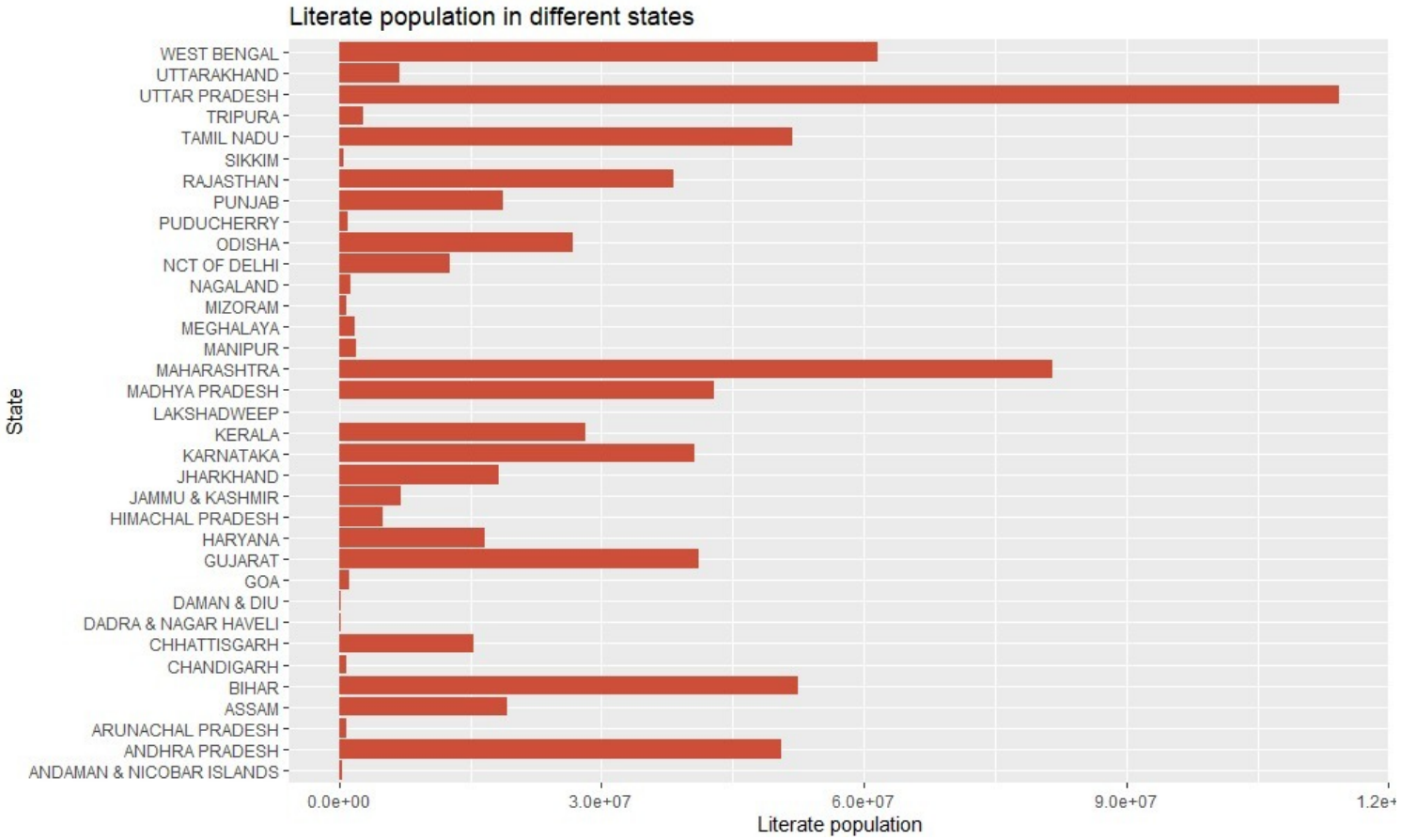
	Name	P_LIT
4	JAMMU & KASHMIR	7067233
7	HIMACHAL PRADESH	5039736
10	PUNJAB	18707137
13	CHANDIGARH	805438
16	UTTARAKHAND	6880953
19	HARYANA	16598988
22	NCT OF DELHI	12737767
25	RAJASTHAN	38275282
28	UTTAR PRADESH	114397555
31	BIHAR	52504553
34	SIKKIM	444952
37	ARUNACHAL PRADESH	766005
40	NAGALAND	1342434
43	MANIPUR	1908476
46	MIZORAM	848175
49	TRIPURA	2804783
52	MEGHALAYA	1785005
55	ASSAM	19177977
58	WEST BENGAL	61538281
61	JHARKHAND	18328069
64	ODISHA	26742595
67	CHHATTISGARH	15379922
70	MADHYA PRADESH	42851169
73	GUJARAT	41093358
76	DAMAN & DIU	188406
79	DADRA & NAGAR HAVELI	223230
82	MAHARASHTRA	81554290
85	ANDHRA PRADESH	50556760
88	KARNATAKA	40647322
91	GOA	1165487
94	LAKSHADWEEP	52553
97	KERALA	28135824
100	TAMIL NADU	51837507
103	PUDUCHERRY	957309
106	ANDAMAN & NICOBAR ISLANDS	294281

A barplot of the data literate_state can be plotted using the code:-

literacy_count= ggplot(literate_state, aes(P_LIT,Name))+geom_bar(stat="identity",fill="tomato3")+ggtitle("Literate population in different states")+xlab('Literate population')+ylab('State')

literacy_count

Output:-



From the data and the graph, it can be seen that the most literate state is Uttar Pradesh and the least literate state is Lakshadweep (results are because of the population difference between the states).

Review 2:-

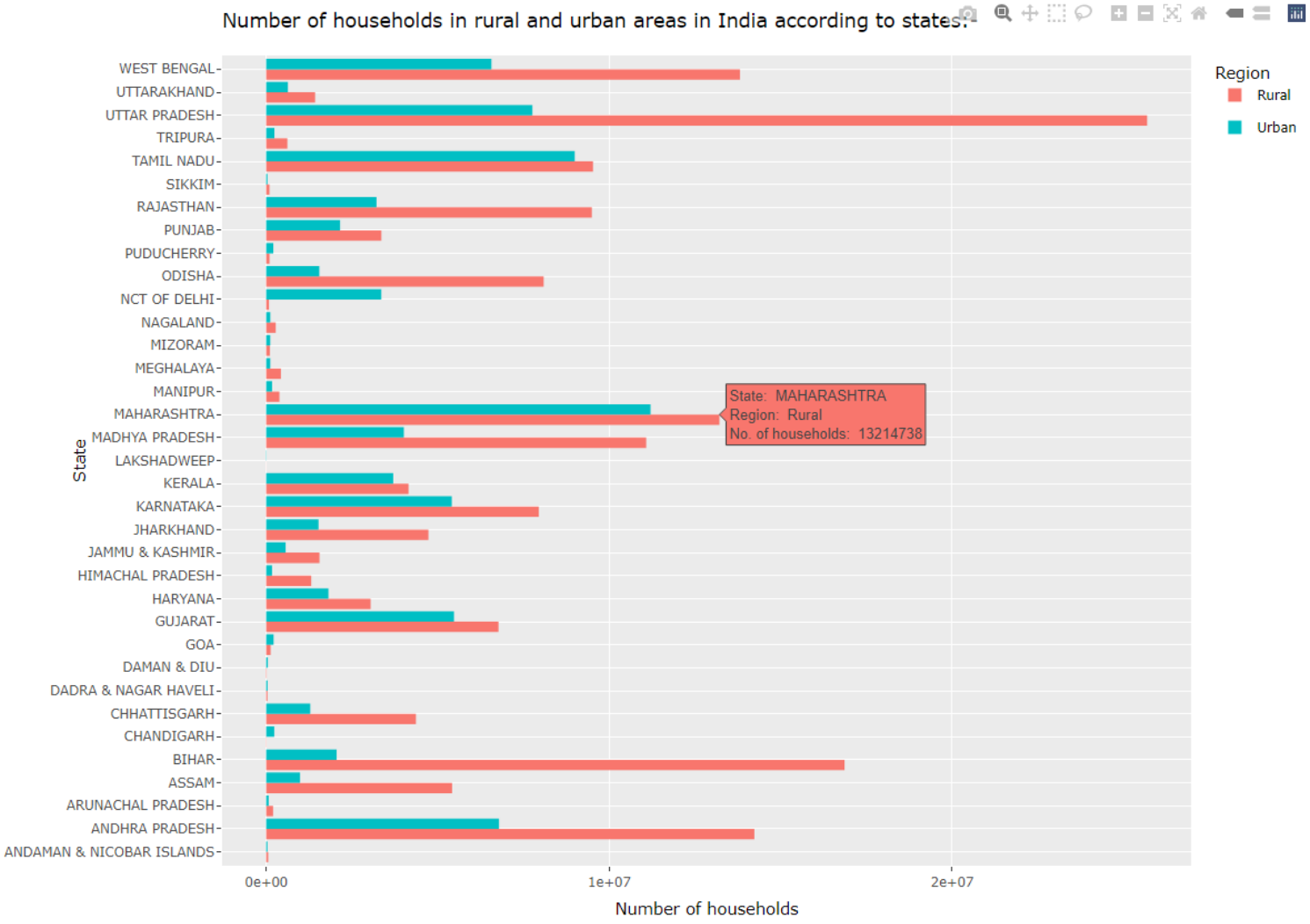
Interactive Plots:-

1. Interactive bar plot for no. of households in rural and urban areas state wise

Code:-

```
HH_RU=subset(r,Level=="STATE"&(TRU=="Rural"|TRU=="Urban"),select=c(Name,TRU,No_HH))

HH_RU_plot=ggplot(HH_RU,aes(No_HH,Name,text=paste("State: ",Name,"<br>Region:
",TRU,"<br>No. of households:
",No_HH)))+geom_bar(aes(fill=TRU),position="dodge",stat="identity")+xlab("Number of
households")+ylab("State")+ggtitle("Number of households in rural and urban areas in India
according to states:-")+labs(fill="Region")

ggplotly(HH_RU_plot, tooltip = "text")
```

Output:-



Interpretation:-

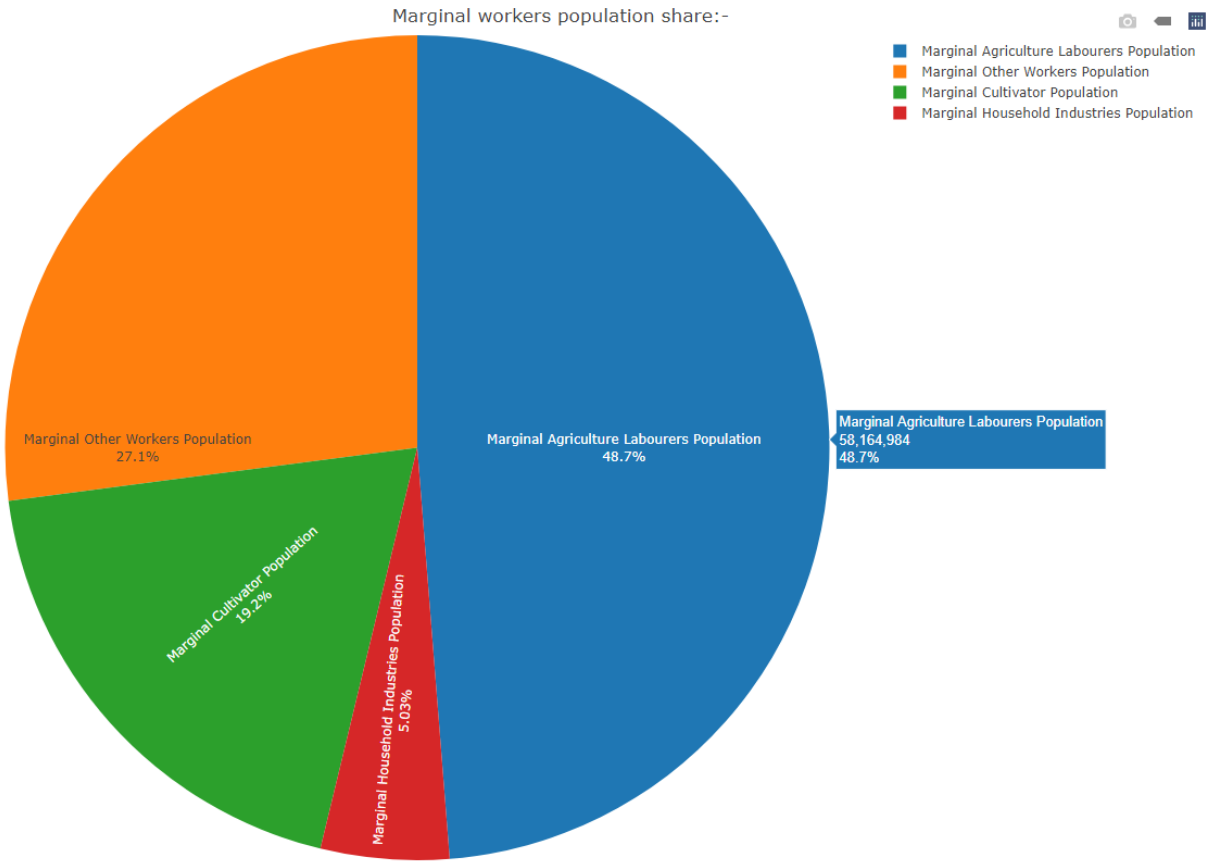
The obtained plot is a grouped bar chart having the number of households on x-axis and the name of states on the y-axis. We selected a grouped bar chart since we had to show data for 2 types of regions of India, i.e., Urban and Rural. The fill color is based on whether the region is rural and urban and the desired region can be displayed by selecting the region using the legend. On hovering on the bars, the state name, region and number of households are displayed.

2. marginal working population interactive pie chart

Code:-

```
marg_work_share=subset(r,Level=="India"&TRU=="Total",select=c(MARG_CL_P,MARG_AL_P,MARG_HH_P,MARG_OT_P))

marg_work_share2=data.frame(t(marg_work_share))

colnames(marg_work_share2)=c("Values")

marg_work_share_plot=plot_ly(data=marg_work_share2,labels=c("Marginal Cultivator Population","Marginal Agriculture Labourers Population","Marginal Household Industries Population","Marginal Other Workers Population"),values=~Values,type="pie",textinfo="label+percent",insidetextorientation = "radial")%>%layout(title="Marginal workers population share:-")

marg_work_share_plot
```

Output:-



Interpretation:-

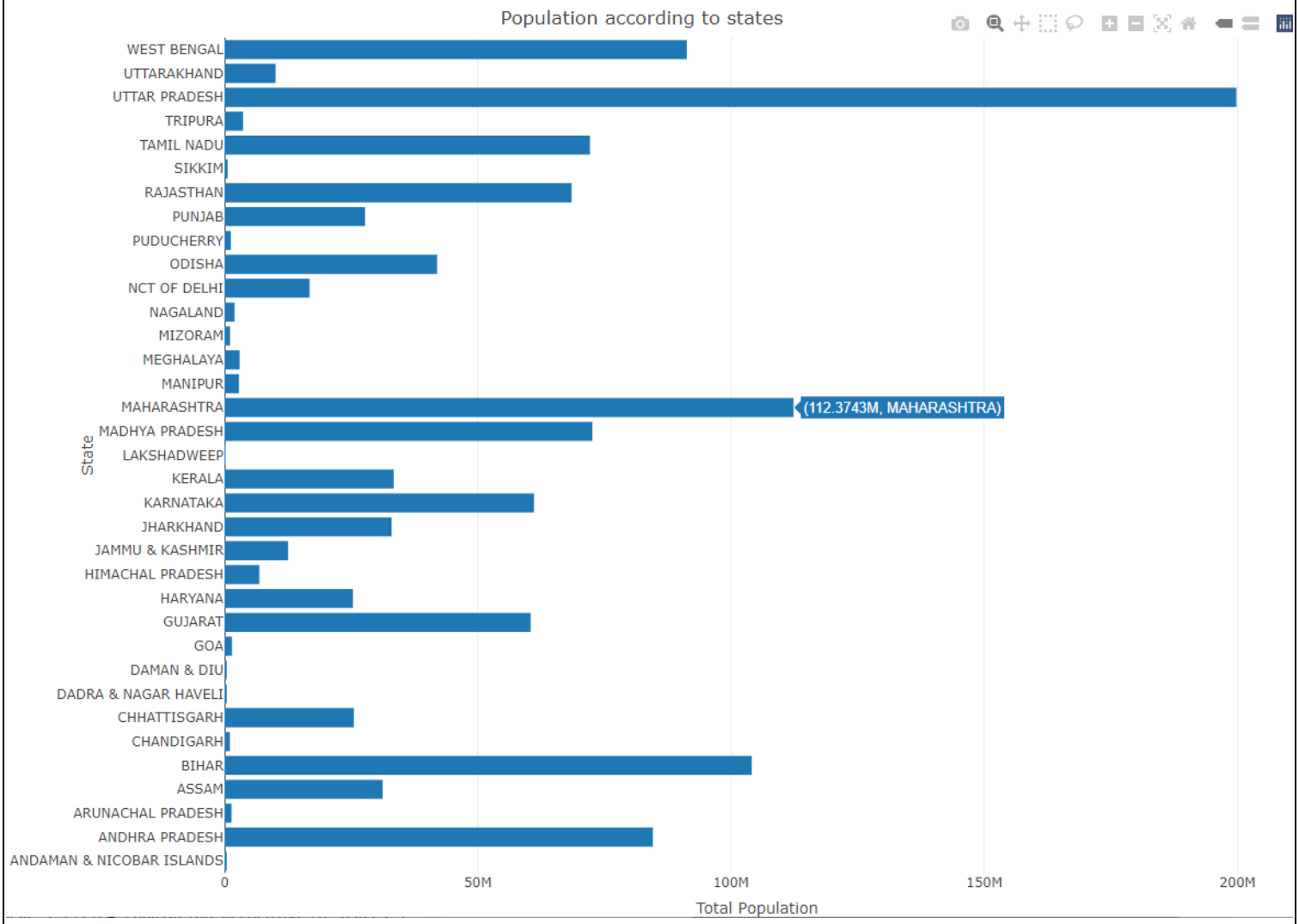
The obtained plot is an interactive pie chart which represents the share of Marginal working population of our country according to the 2011 census data. The percentage of each portion is displayed along with the name. On hovering on each part, the text shown is the name of type of marginal population, the population and its percentage. The parts of pie to be displayed can be selected using the legend.

3. Total population of each state according to 2011 census data.

Code:-

```
pop_state=subset(r,Level=="STATE"&TRU=="Total",select=c(TOT_P,Name))

pop_state_graph=plot_ly(data=pop_state,x=~TOT_P,y=~Name,type='bar')%>%layout(xaxis=list(
title='Total Population'),yaxis=list(title='State',tickmode='linear'),title='Population according to
states')

pop_state_graph
```

Output:-



Interpretation:-

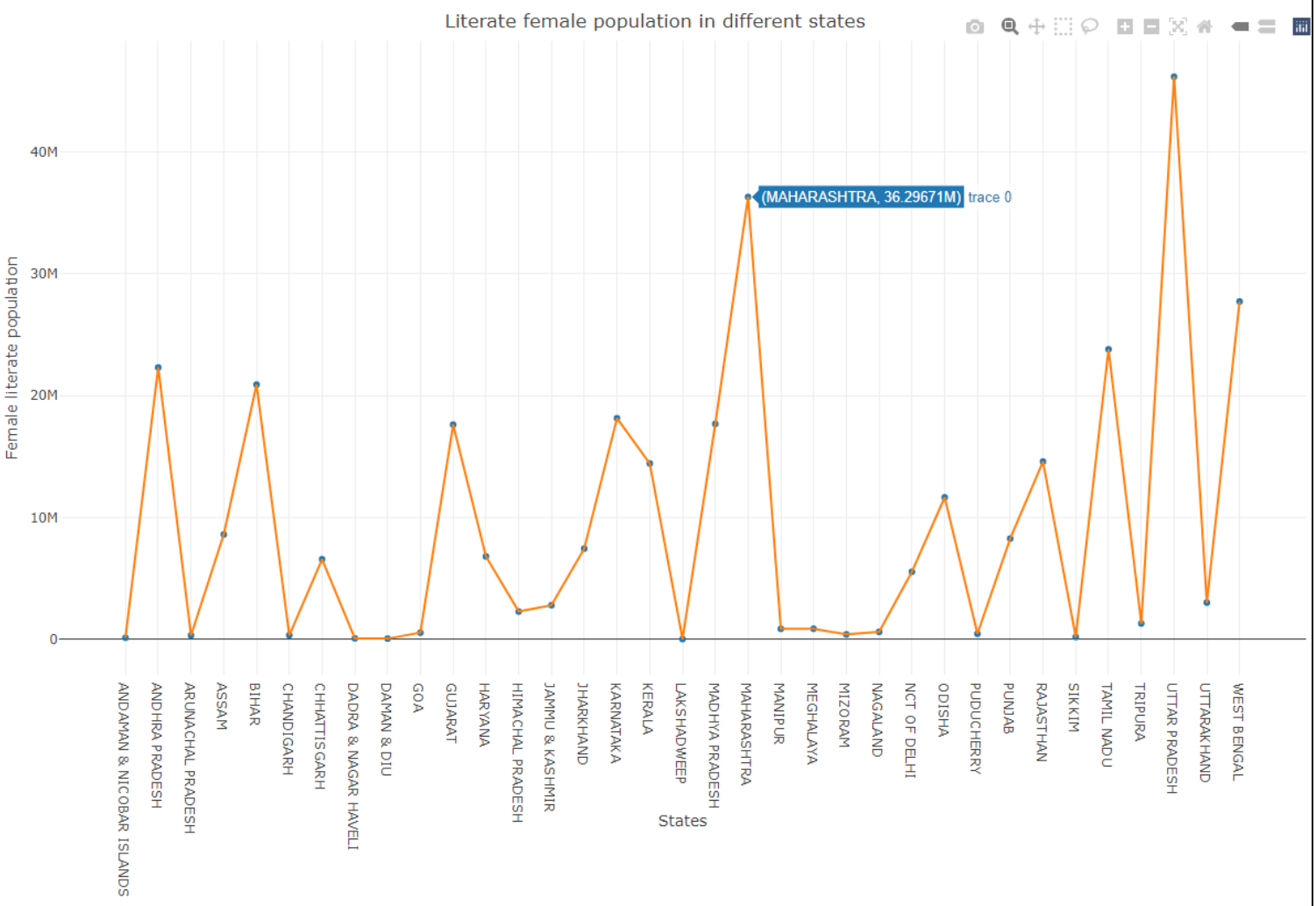
The obtained plot is a simple bar chart showing the total population of each state on the x-axis and the name of each state on the y-axis. Since, we had categorical vs quantitative data for this one, we selected a bar chart. On hovering on the bars, (population, state name) is displayed.

4. state wise female literacy rate (line)

Code:-

```
f_lit=subset(r,Level=="STATE"&TRU=="Total",select=c(Name,F_LIT))

f_lit_plot=plot_ly(f_lit,x=~Name,y=~F_LIT,type='scatter',mode='markers')%>%layout(title='Liter
ate female population in different states',xaxis=list(title='States'),yaxis=list(title='Female literate
population'),showlegend=FALSE)

add_lines(f_lit_plot)
```

Output:-



Interpretation:-

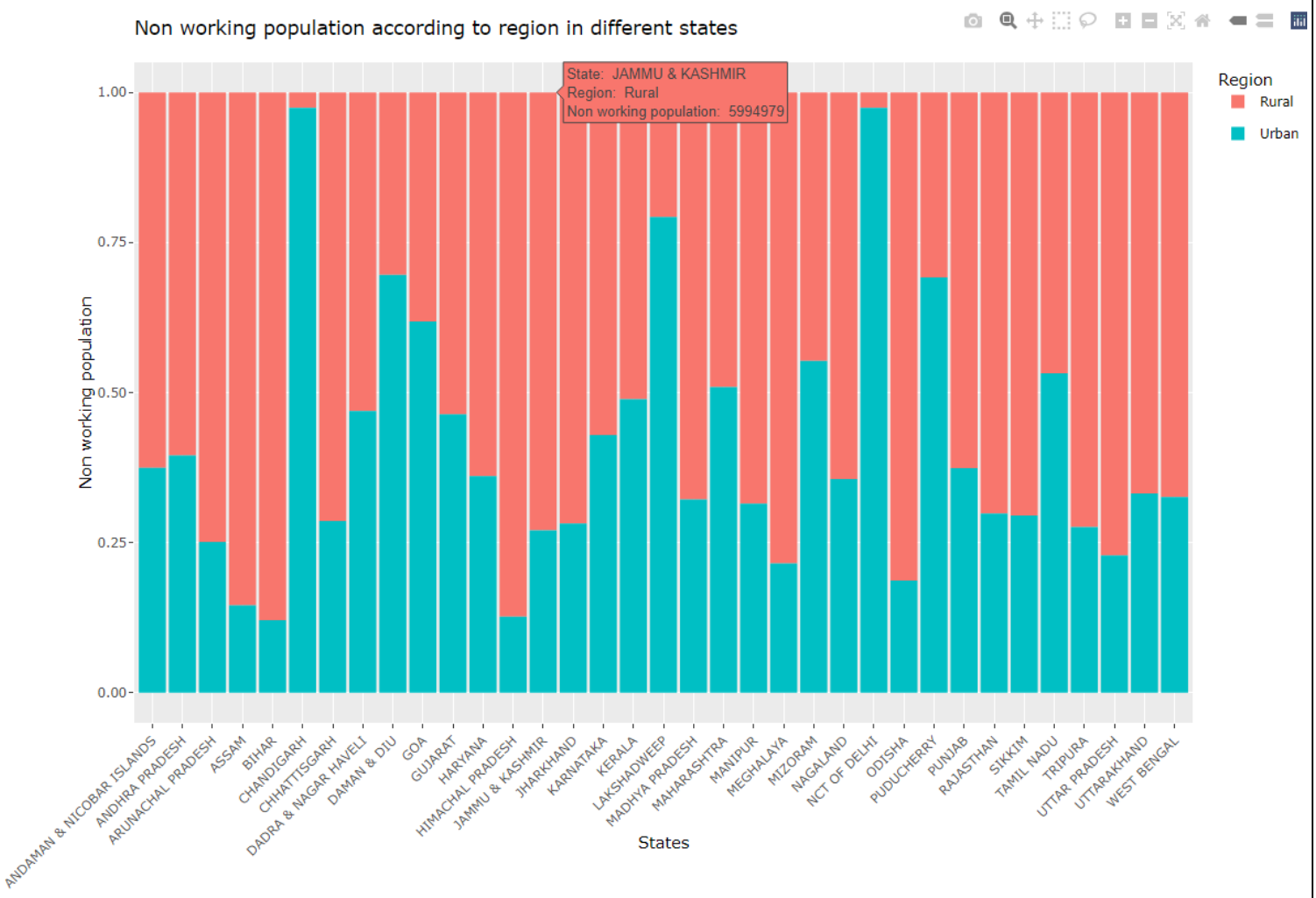
The output shows a line graph with the state names (in alphabetical order) on the x-axis and count of literate female population on the y-axis. We also obtained the scatter points to be displayed on the line so as to clearly show the population values. On hovering on the points, (state name, literate female count) are displayed. Although this is a categorical vs quantitative dataset, line plot is used since it makes this visualization look pleasing.

5. state wise non working population in urban and rural areas (segmented)

Code:-

```
nonwork_RU=subset(r,Level=="STATE"&(TRU=="Rural"|TRU=="Urban"),select=c(Name,TRU,NON_WORK_P))

nonwork_RU_plot=ggplot(nonwork_RU,aes(x=Name,y=NON_WORK_P,fill=TRU,text=paste("State: ",Name,"<br>Region: ",TRU,"<br>Non working population: ",NON_WORK_P)))+geom_bar(stat="identity",position="fill")+xlab("States")+ylab("Non working population")+ggtitle("Non working population according to region in different states")+theme(axis.text.x=element_text(angle=45,hjust=1))+labs(fill="Region")

ggplotly(nonwork_RU_plot, tooltip = "text")
```

Output:-



Interpretation:-

The above plot is a segmented bar chart which has state names as x-axis and non-working population fraction for rural and urban areas as y-axis. Since the data comprises of one quantitative and one categorical attribute, which is further divided into 2 more categories, we used a segmented bar chart for showing how the non-working population is divided between rural and urban areas in different states of India. The fill color used is based on whether the region is rural or urban. On hovering on the bars, the state name, region and the non-working population count are displayed.

Review 3:-

Dashboard Code:-

```
library(shiny)

require(shinydashboard)

library(ggplot2)

library(plotly)

library(dplyr)


r=read.csv("W:/Study
Material/DV/Project/Primary_Census_Abstract_Total_Table_For_India.csv")

pop_state=subset(r,Level=="STATE"&TRU=="Total",select=c(TOT_P,Name))

sex_ratio=subset(r,Level=="STATE"&TRU=="Total",select=c(Name,TOT_M,TOT_F))

tru_work=subset(r,Level=="India",select=c(TRU,TOT_WORK_M,TOT_WORK_F))

main_work_share=
subset(r,Level=="India"&TRU=="Total",select=c(MAIN_CL_P,MAIN_AL_P,MAIN_HH_P,MAIN_OT
_P))

work_share_plot_data=c(main_work_share$MAIN_CL_P,main_work_share$MAIN_AL_P,main_
work_share$MAIN_HH_P,main_work_share$MAIN_OT_P)

piepercent=round(100*work_share_plot_data/sum(work_share_plot_data),1)

nonwork=subset(r,Level=="India",select=c(TRU,NON_WORK_P,NON_WORK_M,NON_WORK_F))

tru = subset(r,Level=="India",select=c(TOT_P,TRU))

r11 <-  c(833748852,377106125)

labelsr11 <-  c("Rural", "Urban")

piepercent<- round(100 * r11 / 1210854977, 1)

literate_state=subset(r,Level=="STATE"&TRU=="Total",select=c(Name,P_LIT))
```

```
HH_RU=subset(r,Level=="STATE"&(TRU=="Rural"|TRU=="Urban"),select=c(Name,TRU,No_HH))

marg_work_share=subset(r,Level=="India"&TRU=="Total",select=c(MARG_CL_P,MARG_AL_P,MARG_HH_P,MARG_OT_P))

marg_work_share2=data.frame(t(marg_work_share))

f_lit=subset(r,Level=="STATE"&TRU=="Total",select=c(Name,F_LIT))

nonwork_RU=subset(r,Level=="STATE"&(TRU=="Rural"|TRU=="Urban"),select=c(Name,TRU,NON_WORK_P))




header <- dashboardHeader(title = "2011 Indian Census Analytics")


sidebar <- dashboardSidebar(

  sidebarMenu(

    menuItem("Non-interactive Plots", tabName = "review_1", icon = icon("signal",lib='glyphicon')),

    menuItem("Interactive Plots", tabName = "review_2", icon = icon("stats",lib='glyphicon')),

    menuItem("About", tabName = "about", icon = icon("info-sign",lib='glyphicon'))

  )

)




frow1 <- fluidRow(
```

```
  box(
    title = "Population according to states"
    ,status = "primary"
    ,solidHeader = TRUE
    ,collapsible = TRUE
    ,plotOutput("pop_state", height = "400px")
  )

  ,box(
    title = "Male and female count"
    ,status = "primary"
    ,solidHeader = TRUE
    ,collapsible = TRUE
    ,plotOutput("sex_ratio", height = "400px")
  )

)

frow2 <- fluidRow(

  box(
    title = "Total male working population"
    ,status = "primary"
    ,solidHeader = TRUE
```

```
    ,collapsible = TRUE
    ,plotOutput("tru_work", height = "400px")
  )

  ,box(
    title = "Total female working population"
    ,status = "primary"
    ,solidHeader = TRUE
    ,collapsible = TRUE
    ,plotOutput("tru_work_1", height = "400px")
  )
)

frow3 <- fluidRow(

  box(
    title = "Main workers population share"
    ,status = "primary"
    ,solidHeader = TRUE
    ,collapsible = TRUE
    ,plotOutput("main_work_share", height = "400px")
  )

  ,box(
```

```
    title = "Unemployed population in different types of areas"
    ,status = "primary"
    ,solidHeader = TRUE
    ,collapsible = TRUE
    ,plotOutput("nonwork", height = "400px")
  )
)

frow4 <- fluidRow(

  box(
    title = "Rural and urban population ratio"
    ,status = "primary"
    ,solidHeader = TRUE
    ,collapsible = TRUE
    ,plotOutput("tru", height = "400px")
  )

  ,box(
    title = "Literate population in different states"
    ,status = "primary"
    ,solidHeader = TRUE
    ,collapsible = TRUE
    ,plotOutput("literate_state", height = "400px")
```

```
  )
)

frow5 <- fluidRow(

  box(
    title = "Households in states at different levels"
    ,status = "primary"
    ,solidHeader = TRUE
    ,collapsible = TRUE
    ,plotlyOutput("HH_RU", height = "550px")
  )

  ,box(
    title = "Literate female population in different states"
    ,status = "primary"
    ,solidHeader = TRUE
    ,collapsible = TRUE
    ,plotlyOutput("f_lit", height = "550px")
  )
)

frow6 <- fluidRow(
```

```
  box(
    title = "Non working population according to region in different states"
    ,status = "primary"
    ,solidHeader = TRUE
    ,collapsible = TRUE
    ,plotlyOutput("nonwork_RU", height = "500px")
  )

  ,box(
    title = "Population according to states"
    ,status = "primary"
    ,solidHeader = TRUE
    ,collapsible = TRUE
    ,plotlyOutput("pop_state_int", height = "500px")
  )
)

frow7 <- fluidRow(

  box(
    title = "Marginal workers population share"
    ,status = "primary"
    ,solidHeader = TRUE
    ,collapsible = TRUE
```

```
    ,plotlyOutput("marg_work_share", height = "400px", width="950px")
  )
)



body <- dashboardBody(
  tabItems(
    tabItem(
      "review_1",frow1,frow2,frow3,frow4
    ),
    tabItem(
      "review_2",frow5,frow6,frow7
    ),
    tabItem(
      "about",
      fluidPage(
        h1("This project has been made by:-"),
        h3("Arush Saxena"),
        br(),
        h1("Acknowledgement:-"),
        p("I am very thankful to my guiding professor, Lydia Jane G. for providing me with the opportunity to work on this project through which I learnt something very functional."),
        p("I got the opportunity to explore many new aspects of the concerned subject throughout the duration of this project.")
      )
```

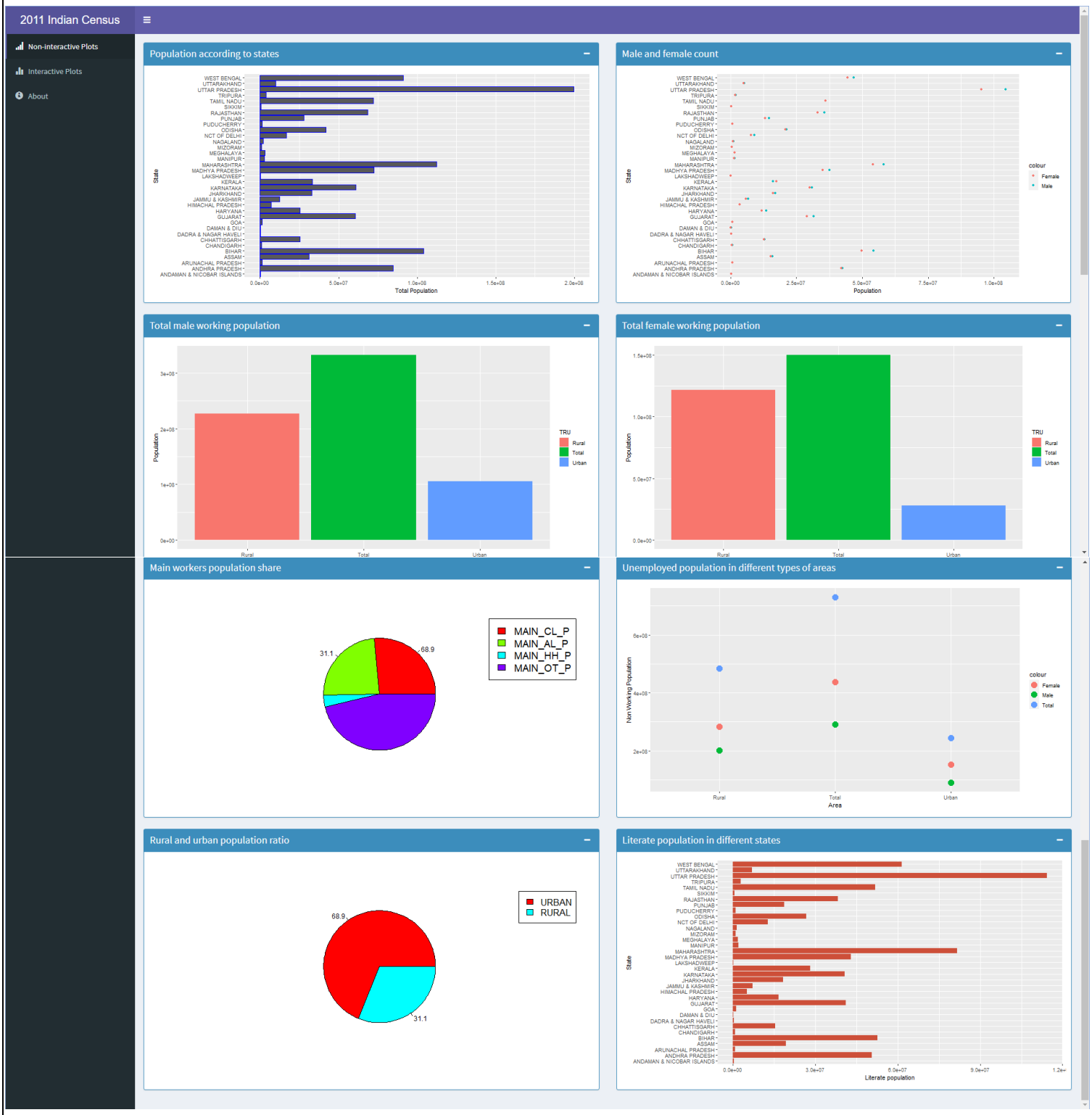
```
    )
  )

)

ui <- dashboardPage(title = 'CSE3020 Project Review 3', header, sidebar, body, skin='purple')

server <- function(input, output) {

  output$pop_state <- renderPlot({

ggplot(pop_state,aes(x=TOT_P,y=Name))+geom_bar(stat="identity",colour="blue")+xlab("Total Population")+ylab("State")
  })

  output$sex_ratio <- renderPlot({

ggplot(sex_ratio)+geom_point(aes(TOT_M,Name,col='Male'))+geom_point(aes(TOT_F,Name,col='Female'))+xlab('Population')+ylab('State')
  })

  output$tru_work <- renderPlot({
    ggplot(tru_work, aes(TRU,TOT_WORK_M,
fill=TRU))+geom_bar(stat="identity")+xlab("Area")+ylab("Population")
  })
```

```
  output$tru_work_1 <- renderPlot({

    ggplot(tru_work, aes(TRU,TOT_WORK_F,
fill=TRU))+geom_bar(stat="identity")+xlab("Area")+ylab("Population")

  })


  output$main_work_share <- renderPlot({

    pie(work_share_plot_data,labels=piepercent,col=rainbow(length(work_share_plot_data)))


legend("topright",c("MAIN_CL_P","MAIN_AL_P","MAIN_HH_P","MAIN_OT_P"),cex=1.5,fill=rain
bow(length(work_share_plot_data)))

  })


  output$nonwork <- renderPlot({


ggplot(nonwork)+geom_point(size=5,aes(TRU,NON_WORK_P,col="Total"))+geom_point(size=5,
aes(TRU,NON_WORK_M,col="Male"))+geom_point(size=5,aes(TRU,NON_WORK_F,col="Female
"))+xlab("Area")+ylab("Non Working Population")

  })


  output$tru <- renderPlot({

    pie(r11, labels = piepercent, col = rainbow(length(r11)))

    legend("topright", c("URBAN", "RURAL"),cex = 1.3, fill = rainbow(length(r11)))

  })


  output$literate_state <- renderPlot({
```

```
   ggplot(literate_state,
aes(P_LIT,Name))+geom_bar(stat="identity",fill="tomato3")+xlab('Literate
population')+ylab('State')

  })


  output$HH_RU <- renderPlotly({

   HH_RU_plot=ggplot(HH_RU,aes(No_HH,Name,text=paste("State: ",Name,"<br>Region:
",TRU,"<br>No. of households:
",No_HH)))+geom_bar(aes(fill=TRU),position="dodge",stat="identity")+xlab("Number of
households")+ylab("State")+labs(fill="Region")

   ggplotly(HH_RU_plot, tooltip = "text")

  })


  output$marg_work_share <- renderPlotly({

   colnames(marg_work_share2)=c("Values")

   plot_ly(data=marg_work_share2,labels=c("Marginal Cultivator Population","Marginal
Agriculture Labourers Population","Marginal Household Industries Population","Marginal Other
Workers
Population"),values=~Values,type="pie",textinfo="label+percent",insidetextorientation =
"radial")

  })


  output$f_lit <- renderPlotly({


f_lit_plot=plot_ly(f_lit,x=~Name,y=~F_LIT,type='scatter',mode='markers')%>%layout(xaxis=list(t
itle='States'),yaxis=list(title='Female literate population'),showlegend=FALSE)

   add_lines(f_lit_plot)

  })
```

```
  output$nonwork_RU <- renderPlotly({


nonwork_RU_plot=ggplot(nonwork_RU,aes(x=Name,y=NON_WORK_P,fill=TRU,text=paste("Stat
e: ",Name,"<br>Region: ",TRU,"<br>Non working population:
",NON_WORK_P)))+geom_bar(stat="identity",position="fill")+xlab("States")+ylab("Non working
population")+theme(axis.text.x=element_text(angle=45,hjust=1))+labs(fill="Region")

    ggplotly(nonwork_RU_plot, tooltip = "text")

  })



  output$pop_state_int <- renderPlotly({

    plot_ly(data=pop_state,x=~TOT_P,y=~Name,type='bar')%>%layout(xaxis=list(title='Total
Population'),yaxis=list(title='State',tickmode='linear'))

  })



}


shinyApp(ui, server)
```

Output:-


2011 Indian Census
Non-interactive Plots
Interactive Plots
About
Population according to states
Total Population
Male and female count
Population
colour
Female
Male
Total male working population
Total female working population
Rural
Total
Urban
TRU
Main workers population share
MAIN_CL_P
MAIN_AL_P
MAIN_HH_P
MAIN_OT_P
68.9
31.1
Unemployed population in different types of areas
Non Working Population
Area
Rural and urban population ratio
URBAN
RURAL
Literate population in different states
Literate population

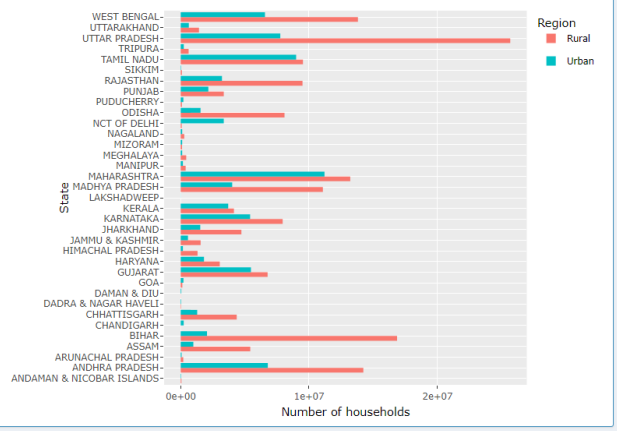
2011 Indian Census

Non-interactive Plots

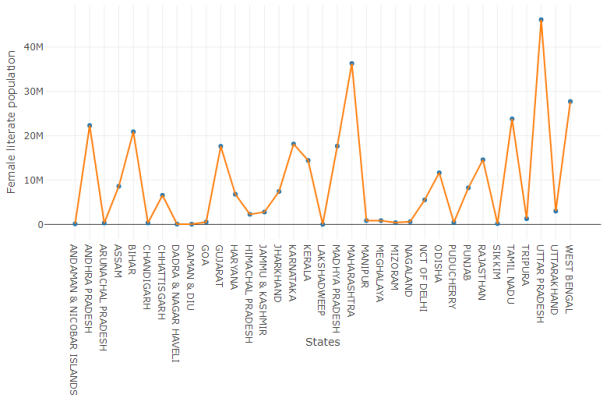
Interactive Plots

About

Households in states at different levels

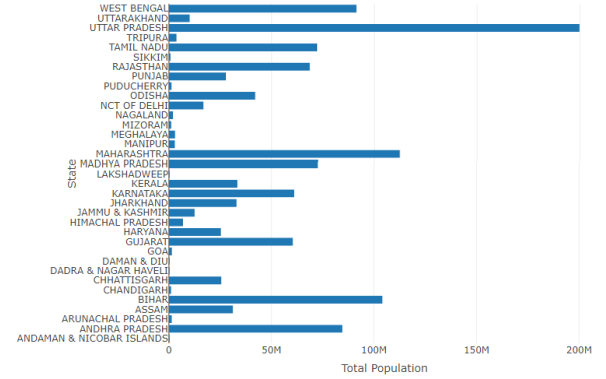


Literate female population in different states

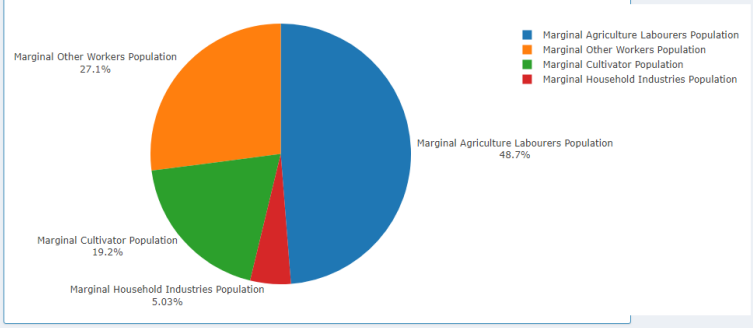


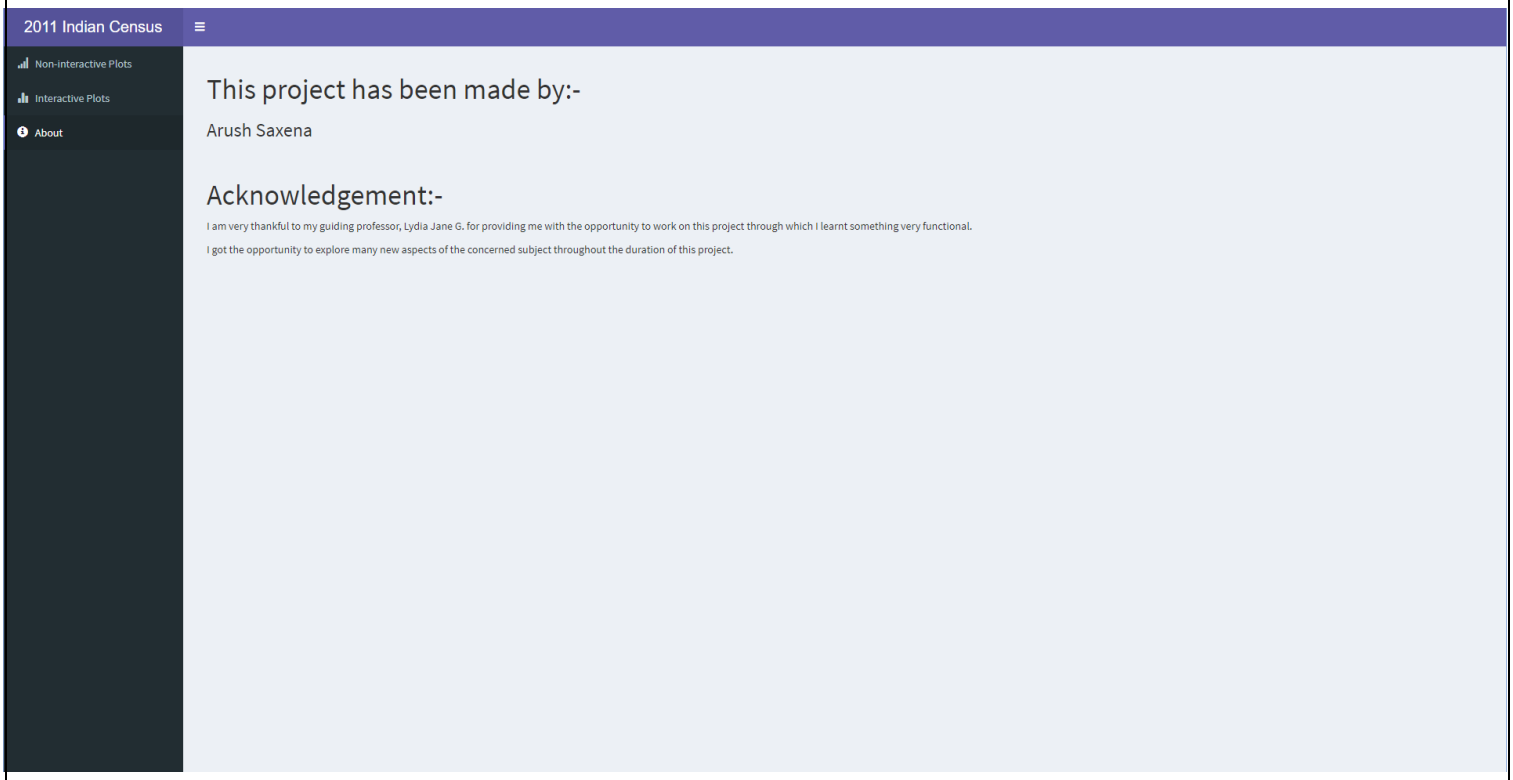
Non working population according to region in different states

Population according to states



Marginal workers population share





Interpretation:-

For the third review of the project component of Data Visualization, I decided to make a dashboard using shiny dashboard package included in R, which has 3 menu items in the sidebar menu. These menu items display sub-pages which consist of Non-interactive plots, interactive plots and some information. The “Non-interactive Plots” menu item takes us to the sub-page containing 8 plots of review 1. The “Interactive Plots” menu item takes us to the sub-page containing 5 interactive plots of review 2. The “About” menu item takes us to the sub-page consisting of the name of the project builder and Acknowledgement. The third review combines all the efforts throughout the 3 reviews which are displayed in this dashboard.

Link for explanation video:-

https://drive.google.com/file/d/1sykc_Q6x7i4Lj2UN16uPszWQClk8sXTC/view?usp=sharing