



DEPARTMENT OF COMPUTER SCIENCE,  
UNIVERSITY OF DELHI

**Masters of Computer Application  
(2021-2024)**

# **REPORT ON EXPLORATORY DATA ANALYSIS**

Using R Language

**Submitted By:**

Hardik Chawda ( Roll no.: 21 )

Tamanna ( Roll no.: 55 )

Anurag Anand ( Roll no.: 62 )

**Submitted To:**

Dr. Vasudha Bhatnagar

# **DESCRIPTION OF DATA SET**

---

**Dataset :** Population Analysis of Delhi 2001

**Data set file name :** PCA0705\_New\_Delhi-2001

**Source :** [data.gov.in](http://data.gov.in)

## **Description :**

This data refers to the area wise population of New Delhi, it gives us an overview of urban and rural areas, total literate population, total illiterate population, working males and females, caste wise population in rural and urban areas, population working as cultivators, population working in agricultural field and the related data.

**Released Under :** National Data Sharing and Accessibility Policy (NDSAP)

**Contributor :** Ministry of Home Affairs  
Department of Home Registrar General and  
Census Commissioner, India

**Sectors :** Population Enumerator, Census,  
Census and Surveys

# ANALYSIS

```
#libraries
library("ggplot2")
library("astsa")
library("matrixStats")
library("dplyr")

mydata <- read.csv("PCA0705_New_Delhi-2001.csv")

head(mydata)
```

Console Terminal x Jobs x

R 4.1.2 · ~/

	Main.Agricultural.Labourers.Population...Persons	Main.Agricultural.Labourers.Population...Males	
1	71	43	
2	0	0	
3	71	43	
4	29	16	
5	0	0	
6	29	16	
	Main.Agricultural.Labourers.Population...Females	Main.Household.Industries.Population...Persons	
1	28	1007	
2	0	0	
3	28	1007	
4	13	156	
5	0	0	
6	13	156	
	Main.Household.Industries.Population...Males	Main.Household.Industries.Population...Females	
1	778	229	
2	0	0	
3	778	229	
4	136	20	
5	0	0	
6	136	20	
	Main.Other.workers.Population...Persons	Main.Other.workers.Population...Males	Main.other.workers.Population...Females
1	63251	51891	11360
2	0	0	0
3	63251	51891	11360
4	20504	16473	4031
5	0	0	0
6	20504	16473	4031
	Marginal.worker.Population...Persons	Marginal.worker.Population...Males	Marginal.worker.Population...Females
1	3169	2443	726
2	0	0	0
3	3169	2443	726
4	646	509	137
5	0	0	0
6	646	509	137
	Marginal.cultivator.Population...Persons	Marginal.cultivator.Population...Males	
1	9	3	
2	0	0	
3	9	3	
4	1	0	
5	0	0	
6	1	0	
	Marginal.cultivator.Population...Females	Marginal.Agriculture.Labourers.Population...Persons	
1	6	13	
2	0	0	
3	6	13	
4	1	1	
5	0	0	
6	1	1	
	Marginal.Agriculture.Labourers.Population...Males	Marginal.Agriculture.Labourers.Population...Females	
1	7	6	
2	0	0	
3	7	6	
4	1	0	
5	0	0	
6	1	0	
	Marginal.Household.Industries.Population...Persons	Marginal.Household.Industries.Population...Males	
1	66	33	
2	0	0	
3	66	33	
4	13	7	
5	0	0	
6	13	7	
	Marginal.Household.Industries.Population...Females	Marginal.other.workers.Population...Persons	
1	33	3081	
2	0	0	
3	33	3081	
4	6	631	
5	0	0	
6	6	631	
	Marginal.other.workers.Population...Males	Marginal.other.workers.Population...Females	Non.Worker.Population...Persons
1	2400	681	111518
2	0	0	0
3	2400	681	111518
4	501	130	37776
5	0	0	0
6	501	130	37776
	Non.worker.Population...Males	Non.worker.Population...Females	
1	44730	66788	
2	0	0	
3	44730	66788	
4	15417	22359	
5	0	0	
6	15417	22359	

> |

## summary(mydata)

```
> summary(mydata)
   State.Code District.Code Sub.Distt      Town..Village.Code Ward.Code Enumeration.Block.Code Level
Min.    :7      Min.    :5      Min.  :0.000  Min.  :0      Min.  :0      Length:28
1st Qu.:7      1st Qu.:5      1st Qu.:1.000  1st Qu.:0      1st Qu.:0      Class :character
Median :7      Median :5      Median :2.000   Median :40501000  Median :0      Mode  :character
Mean   :7      Mean   :5      Mean   :1.786   Mean   :23143643  Mean   :1.464   Mean   :0
3rd Qu.:7      3rd Qu.:5      3rd Qu.:2.250   3rd Qu.:40501000  3rd Qu.:2.250   3rd Qu.:0
Max.   :7      Max.   :5      Max.   :3.000   Max.   :40503000  Max.   :9.000   Max.   :0
   Name          Total..Rural..Urban No.of.Households Total.Population...Persons Total.Population...Males
Length:28      Length:28      Min.  :0      Min.  :0      Min.  :0
Class :character Class :character 1st Qu.: 1513   1st Qu.: 7136   1st Qu.: 3866
Mode  :character Mode  :character  Median : 4762   Median : 22312  Median :12222
                           Mean   : 8493   Mean   : 38381  Mean   :21419
                           3rd Qu.:12408  3rd Qu.: 59132  3rd Qu.:32562
                           Max.   :39633   Max.   :179112  Max.   :99956
Total.Population...Females Population.Population.in.the.Age.Group.0.6...Persons Population.in.the.Age.Group.0.6...Males
Min.   : 0      Min.   : 0      Min.  : 0.0
1st Qu.: 3270  1st Qu.: 788   1st Qu.: 420.5
Median :10090  Median : 2686  Median :1357.0
Mean   :16962  Mean   : 4606  Mean   :2427.2
3rd Qu.:26570  3rd Qu.: 6340  3rd Qu.:3388.0
Max.   :79156  Max.   :21496  Max.   :11327.0
Population.in.the.Age.Group.0.6.Females Scheduled.Castes.Population...Persons Scheduled.Castes.Population...Males
Min.   : 0.0      Min.   : 0      Min.  : 0
1st Qu.: 367.5  1st Qu.: 1245  1st Qu.: 657
Median :1329.0  Median : 5352  Median : 2866
Mean   :2179.1  Mean   : 8529  Mean   : 4651
3rd Qu.:2952.0  3rd Qu.:12072  3rd Qu.: 6501
Max.   :10169.0  Max.   :39803   Max.   :21706
Scheduled.Castes.Population...Females Scheduled.Tribes.Population...Persons Scheduled.Tribes.Population...Males
Min.   : 0.0      Min.   : 0      Min.  : 0
1st Qu.: 588.2  1st Qu.: 0      1st Qu.: 0
Median :2482.5  Median : 0      Median : 0
Mean   :3877.9  Mean   : 0      Mean   : 0
3rd Qu.:5571.0  3rd Qu.: 0      3rd Qu.: 0
Max.   :18097.0  Max.   : 0      Max.   : 0
Scheduled.Tribes.Population...Females Literates.Population...Persons Literates.Population...Males
Min.   : 0      Min.   : 0      Min.  : 0
1st Qu.: 0      1st Qu.: 5470  1st Qu.: 3193
Median :0      Median : 17020  Median : 9931
Mean   :0      Mean   : 28113  Mean   :16830
3rd Qu.:0      3rd Qu.: 45888  3rd Qu.:26743
Max.   :0      Max.   :131196   Max.   :78540
Literates.Population...Females Illiterate...Persons Illiterate...Males Illiterate...Females
Min.   : 0      Min.   : 0      Min.  : 0.0  Min.  : 0
1st Qu.: 1468  1st Qu.: 1600  1st Qu.: 673.5 1st Qu.: 917
Median :7056  Median : 6612  Median : 2874.5 Median: 3601
Mean   :11283  Mean   :10268  Mean   : 4589.1 Mean : 5679
3rd Qu.:19145  3rd Qu.:13560  3rd Qu.: 6168.2 3rd Qu.: 7425
Max.   :52656  Max.   :47916   Max.   :21416.0 Max.  :26500
Total.worker.Population...Persons Total.worker.Population...Males Total.worker.Population...Females
Min.   : 0      Min.   : 0      Min.  : 0.0
1st Qu.: 2671  1st Qu.: 2014  1st Qu.: 334.8
Median : 8336  Median : 6590  Median :1720.0
Mean   :14484  Mean   :11834  Mean   :2650.3
   3rd Qu.:21356      3rd Qu.:17145      3rd Qu.: 4211.0
   Max.   :67594      Max.   :55226      Max.   :12368.0
Main.worker.Population...Persons Main.worker.Population...Males Main.worker.Population...Females
Min.   : 0      Min.   : 0      Min.  : 0.0
1st Qu.: 2587  1st Qu.: 1965  1st Qu.: 292.5
Median : 8066  Median : 6388  Median :1661.0
Mean   :13805  Mean   :11311  Mean   :2494.7
3rd Qu.:20710  3rd Qu.:16636  3rd Qu.: 4074.0
Max.   :64425  Max.   :52783   Max.   :11642.0
Main.cultivator.Population...Persons Main.Cultivator.Population...Males Main.cultivator.Population...Females
Min.   : 0.00      Min.   : 0.00      Min.  : 0.000
1st Qu.: 1.75      1st Qu.: 1.75      1st Qu.: 0.000
Median : 9.50      Median : 5.00      Median : 2.500
Mean   :20.57      Mean   :15.21      Mean   : 5.357
3rd Qu.:21.50      3rd Qu.:15.00      3rd Qu.:10.000
Max.   :96.00      Max.   :71.00      Max.   :25.000
Main.Agricultural.Labourers.Population...Persons Main.Agricultural.Labourers.Population...Males
Min.   : 0.00      Min.   : 0.000
1st Qu.: 1.00      1st Qu.: 1.000
Median : 5.00      Median : 3.500
Mean   :15.21      Mean   : 9.214
3rd Qu.:28.00      3rd Qu.:15.000
Max.   :71.00      Max.   :43.000
Main.Agricultural.Labourers.Population...Females Main.Household.Industries.Population...Persons
Min.   : 0          Min.   : 0.0
1st Qu.: 0          1st Qu.: 22.5
Median : 2          Median : 126.0
Mean   : 6          Mean   : 215.8
3rd Qu.:13          3rd Qu.: 327.5
Max.   :28          Max.   :1007.0
Main.Household.Industries.Population...Males Main.Household.Industries.Population...Females
Min.   : 0.00      Min.   : 0.00
1st Qu.: 15.75     1st Qu.: 4.25
Median : 98.50     Median : 16.00
Mean   :166.71     Mean   : 49.07
3rd Qu.:251.75     3rd Qu.: 66.00
Max.   :778.00     Max.   :229.00
```

Main.other.Workers.Population...Persons	Main.other.Workers.Population...Males	Main.other.Workers.Population...Females
Min. : 0	Min. : 0	Min. : 0.0
1st Qu.: 2523	1st Qu.: 1943	1st Qu.: 283.5
Median : 7872	Median : 6242	Median : 1622.0
Mean : 13554	Mean : 11120	Mean : 2434.3
3rd Qu.: 20504	3rd Qu.: 16473	3rd Qu.: 4031.0
Max. : 63251	Max. : 51891	Max. : 11360.0
Marginal.worker.Population...Persons	Marginal.worker.Population...Males	Marginal.worker.Population...Females
Min. : 0.00	Min. : 0.0	Min. : 0.00
1st Qu.: 58.75	1st Qu.: 40.5	1st Qu.: 18.25
Median : 394.00	Median : 319.0	Median : 72.00
Mean : 679.07	Mean : 523.5	Mean : 155.57
3rd Qu.: 998.00	3rd Qu.: 790.0	3rd Qu.: 208.00
Max. : 3169.00	Max. : 2443.0	Max. : 726.00
Marginal.cultivator.Population...Persons	Marginal.cultivator.Population...Males	
Min. : 0.000	Min. : 0.0000	
1st Qu.: 0.000	1st Qu.: 0.0000	
Median : 1.000	Median : 0.0000	
Mean : 1.929	Mean : 0.6429	
3rd Qu.: 3.000	3rd Qu.: 1.0000	
Max. : 9.000	Max. : 3.0000	
Marginal.cultivator.Population...Females	Marginal.Agriculture.Labourers.Population...Persons	
Min. : 0.000	Min. : 0.000	
1st Qu.: 0.000	1st Qu.: 0.000	
Median : 1.000	Median : 1.000	
Mean : 1.286	Mean : 2.786	
3rd Qu.: 2.000	3rd Qu.: 4.250	
Max. : 6.000	Max. : 13.000	
Marginal.Agriculture.Labourers.Population...Males	Marginal.Agriculture.Labourers.Population...Females	
Min. : 0.0	Min. : 0.000	
1st Qu.: 0.0	1st Qu.: 0.000	
Median : 1.0	Median : 0.000	
Mean : 1.5	Mean : 1.286	
3rd Qu.: 2.0	3rd Qu.: 0.750	
Max. : 7.0	Max. : 6.000	
Marginal.Household.Industries.Population...Persons	Marginal.Household.Industries.Population...Males	
Min. : 0.00	Min. : 0.000	
1st Qu.: 0.75	1st Qu.: 0.000	
Median : 9.00	Median : 5.500	
Mean : 14.14	Mean : 7.071	
3rd Qu.: 24.50	3rd Qu.: 11.000	
Max. : 66.00	Max. : 33.000	
Marginal.Household.Industries.Population...Females	Marginal.other.workers.Population...Persons	
Min. : 0.000	Min. : 0.00	
1st Qu.: 0.000	1st Qu.: 57.25	
Median : 4.000	Median : 387.00	
Mean : 7.071	Mean : 660.21	
3rd Qu.: 12.000	3rd Qu.: 963.00	
Max. : 33.000	Max. : 3081.00	
Marginal.other.Workers.Population...Males	Marginal.other.workers.Population...Females	Non.Worker.Population...Persons
Min. : 0.00	Min. : 0.0	Min. : 0
1st Qu.: 39.75	1st Qu.: 17.5	1st Qu.: 4675
Median : 307.50	Median : 66.5	Median : 13976
Mean : 514.29	Mean : 145.9	Mean : 23897
3rd Qu.: 769.00	3rd Qu.: 194.0	3rd Qu.: 37776
Max. : 2400.00	Max. : 681.0	Max. : 111518
Non.Worker.Population...Males	Non.Worker.Population...Females	
Min. : 0	Min. : 0	
1st Qu.: 1870	1st Qu.: 2805	
Median : 5632	Median : 8344	
Mean : 9585	Mean : 14312	
3rd Qu.: 15417	3rd Qu.: 22359	
Max. : 44730	Max. : 66788	

> |

# GRAPHS

---

## 1- TOTAL POPULATION IN DELHI

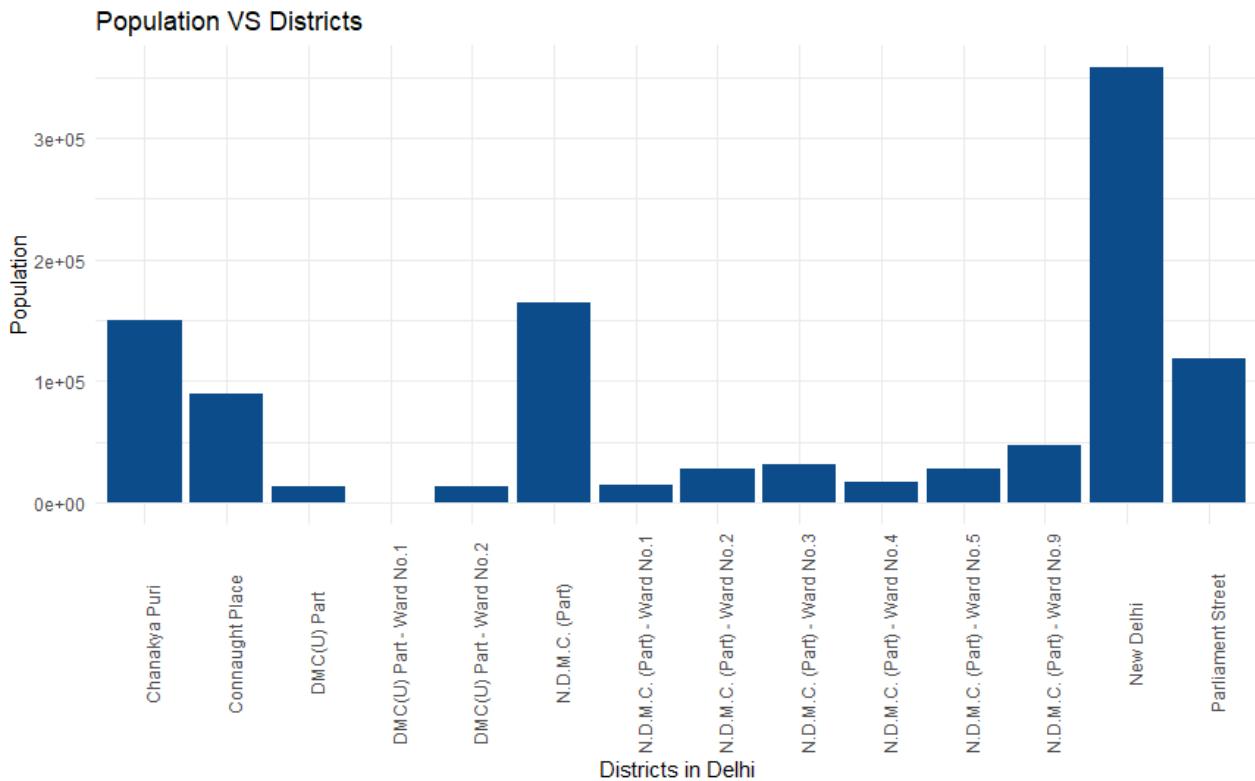
---

```
#libraries used
library("ggplot2")
library("astsa")
library("matrixStats")
library("dplyr")

mydata <- read.csv("PCA0705_New_Delhi-2001.csv")

#graph of population vs district

ggplot(mydata) +
  aes(x = Name, weight = Total.Population...Persons) +
  geom_bar(fill = "#0c4c8a") +
  labs(x = "Districts in Delhi", y = "Population", title = "Population VS Districts")+
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90))
```

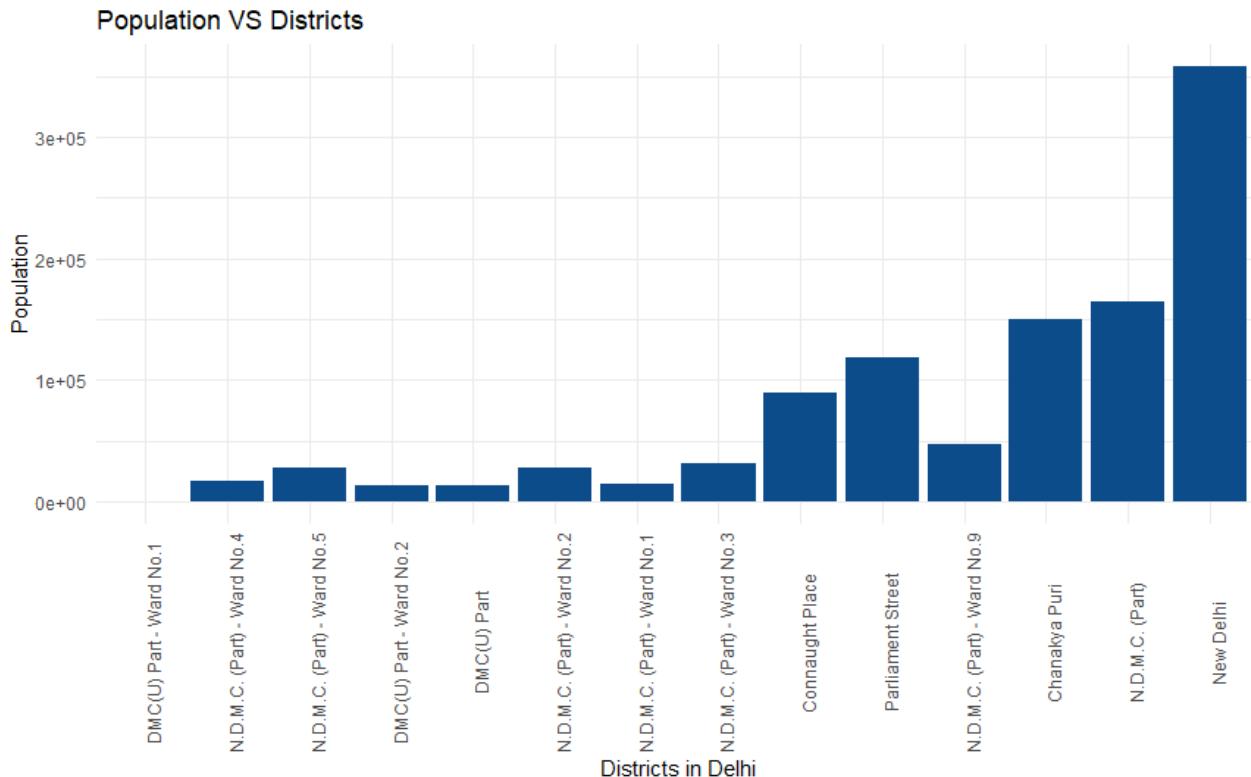


## 2- TOTAL POPULATION IN DELHI (SORTED)

---

```
#graph of population vs district sorted
```

```
ggplot(mydata, aes(reorder(Name,Total.Population...Persons),Total.Population...Persons)) +  
  geom_col(fill = "#0c4c8a") +  
  labs(x = "Districts in Delhi", y = "Population", title = "Population VS  
Districts") + theme_minimal() +  
  theme(axis.text.x = element_text(angle = 90))
```

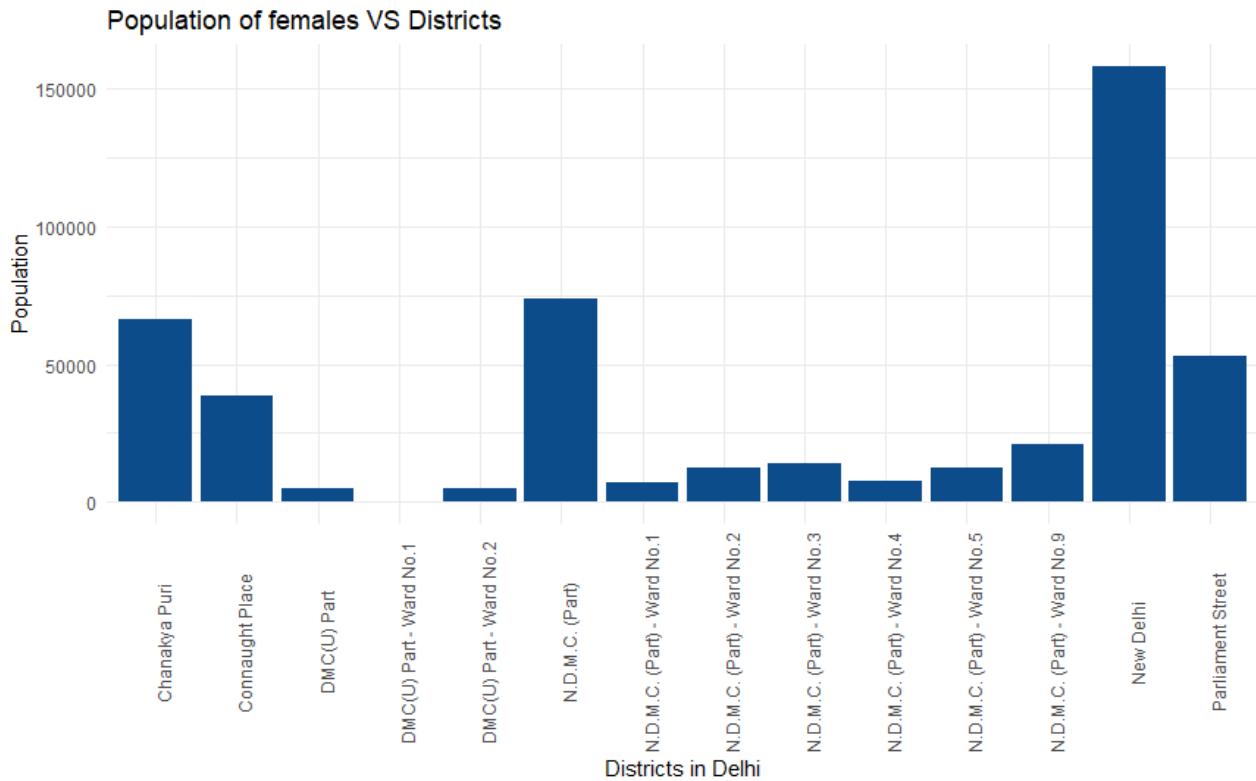


### 3- TOTAL POPULATION OF FEMALES IN DELHI

---

```
#graph of total population of females in each district
```

```
ggplot(mydata) +  
  aes(x = Name, weight = Total.Population...Females) +  
  geom_bar(fill = "#0c4c8a") +  
  labs(x = "Districts in Delhi", y = "Population", title = "Population of females VS  
Districts") + theme_minimal() +  
  theme(axis.text.x = element_text(angle = 90))
```

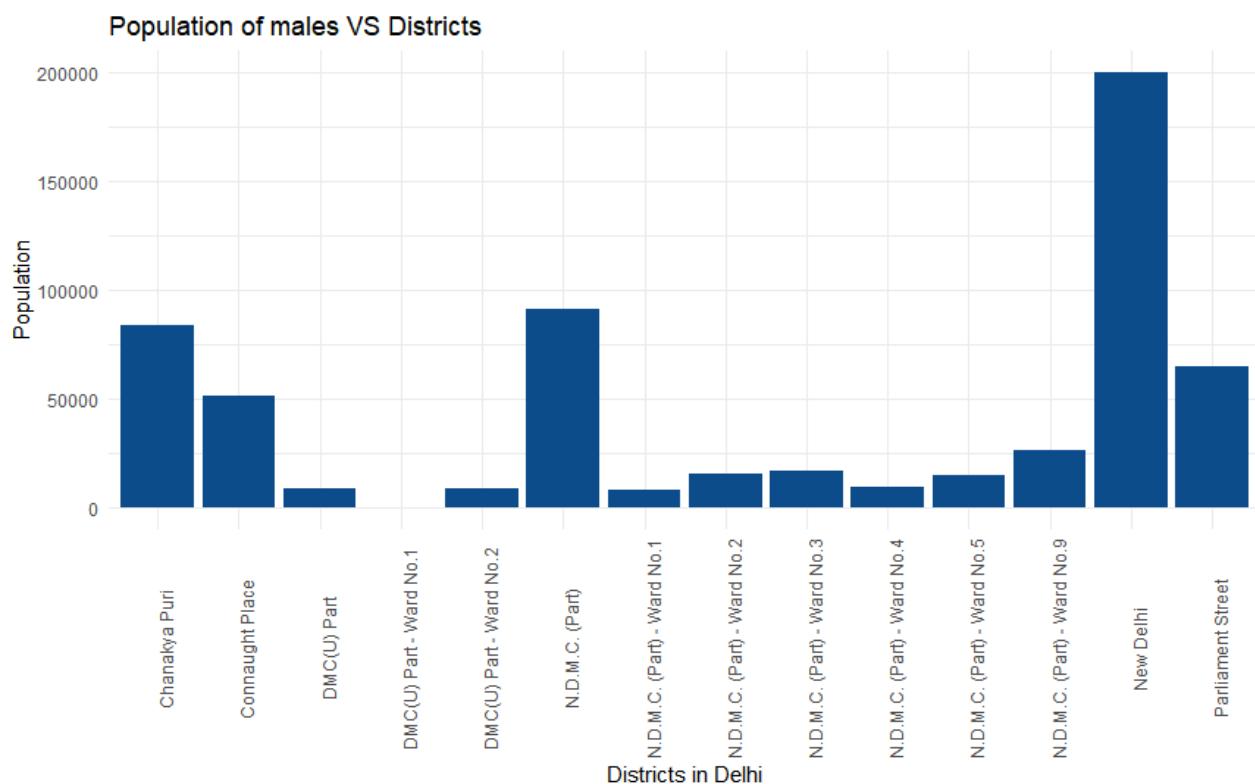


## 4- TOTAL POPULATION OF MALES IN DELHI

---

```
#graph of total population of Males in each district
```

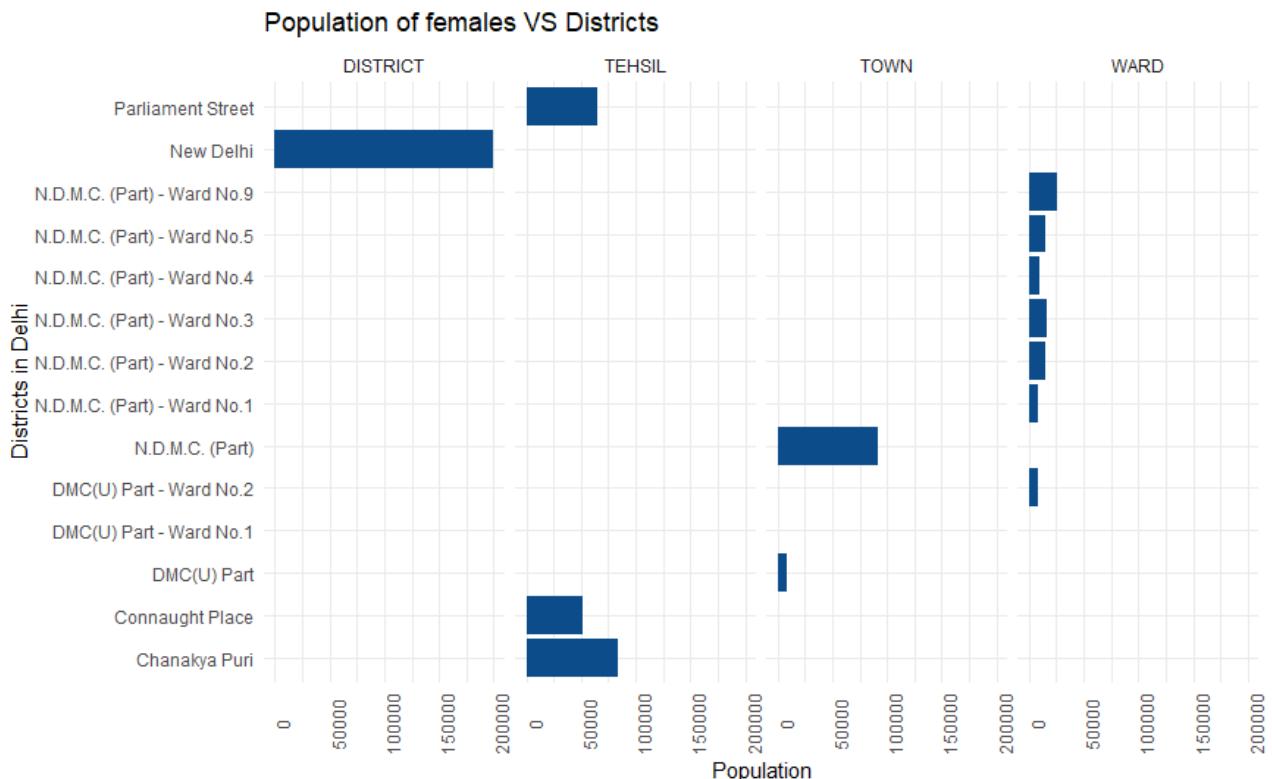
```
ggplot(mydata) +  
  aes(x = Name, weight = Total.Population...Males) +  
  geom_bar(fill = "#0c4c8a") +  
  labs(x = "Districts in Delhi", y = "Population", title = "Population of females VS  
Districts") + theme_minimal() +  
  theme(axis.text.x = element_text(angle = 90))
```



## 5- TOTAL POPULATION OF EACH AREA IN TERMS OF LEVEL

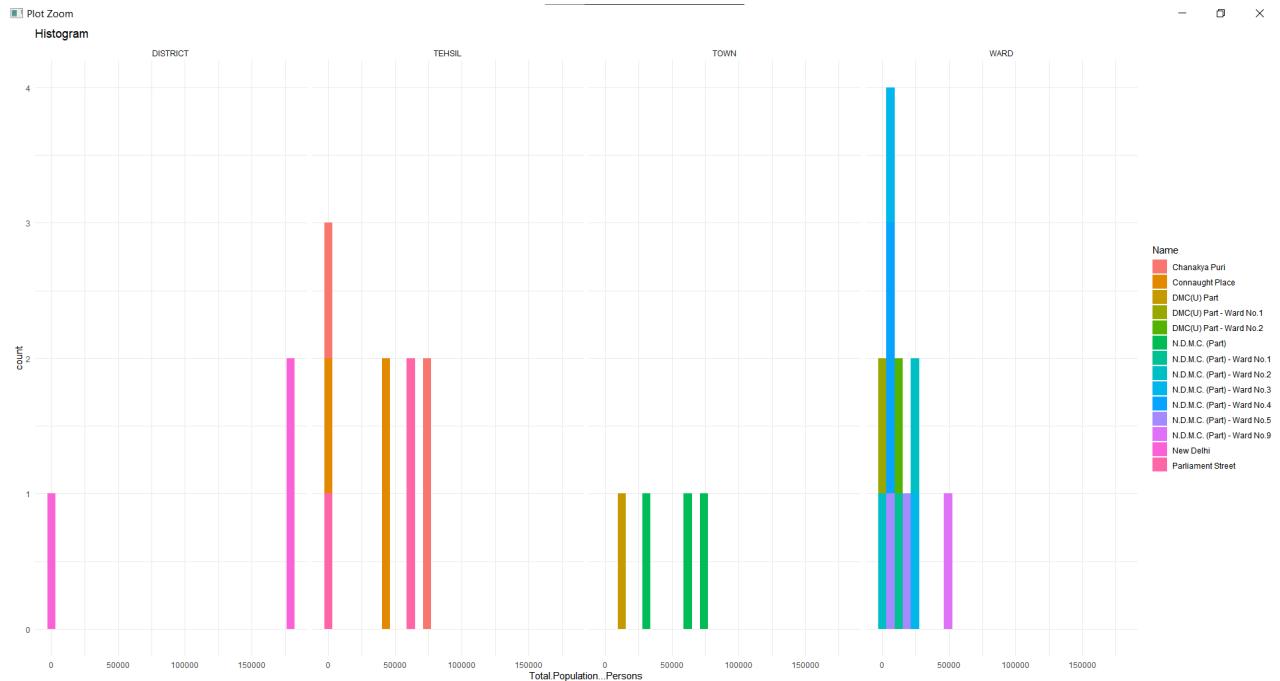
---

```
#graph of total population vs district in terms of levels  
  
ggplot(mydata) +  
  aes(x = Name, weight = Total.Population...Males) +  
  geom_bar(fill = "#0c4c8a") + coord_flip() +  
  labs(x = "Districts in Delhi", y = "Population", title = "Population of females VS  
Districts") + theme_minimal() +  
  theme(axis.text.x = element_text(angle = 90)) + facet_grid(vars(), vars(Level))
```



# 6- HISTOGRAM FOR TOTAL POPULATION OF EACH AREA IN TERMS OF LEVEL

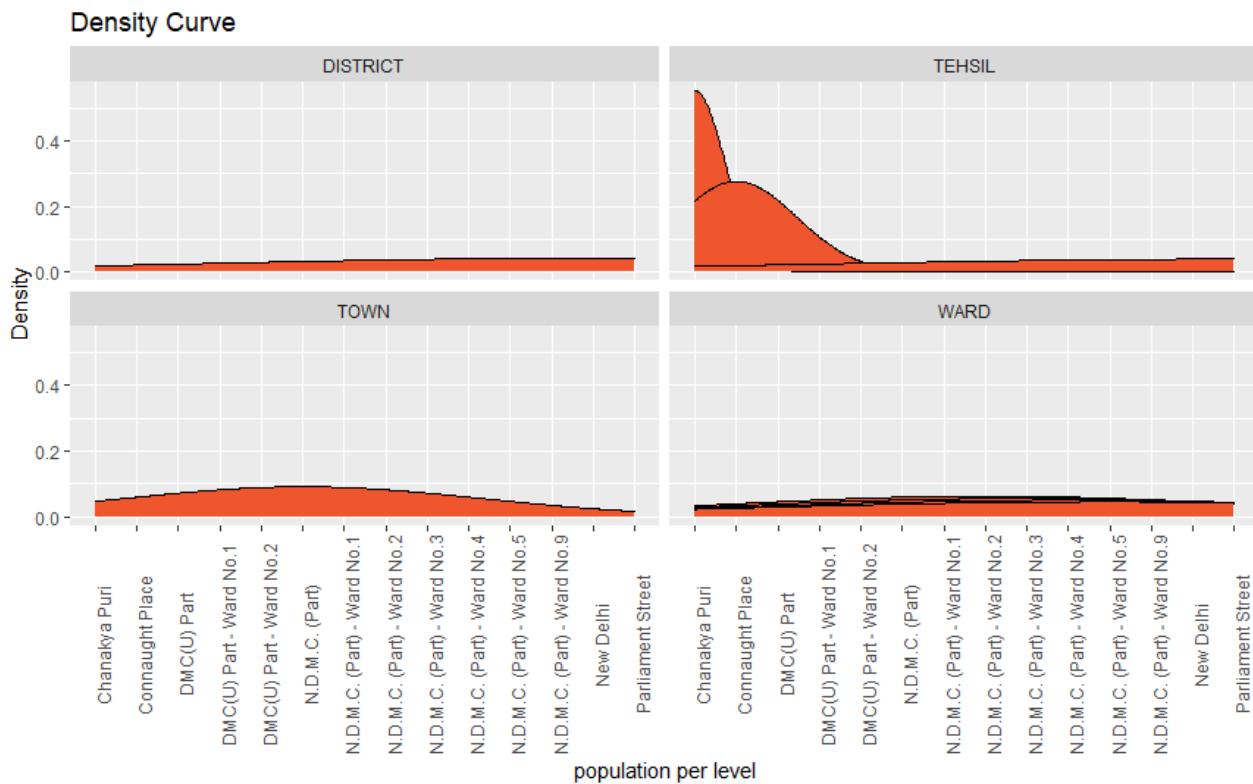
```
#histogram
ggplot(mydata) +
  aes(x = Total.Population...Persons, fill = Name) +
  geom_histogram(bins = 30L) +
  labs(title = "Histogram") +
  scale_fill_hue() +
  theme_minimal() +
  facet_grid(vars(), vars(Level))
```



## 7- POPULATION DENSITY IN EACH LEVEL OF THE STATE

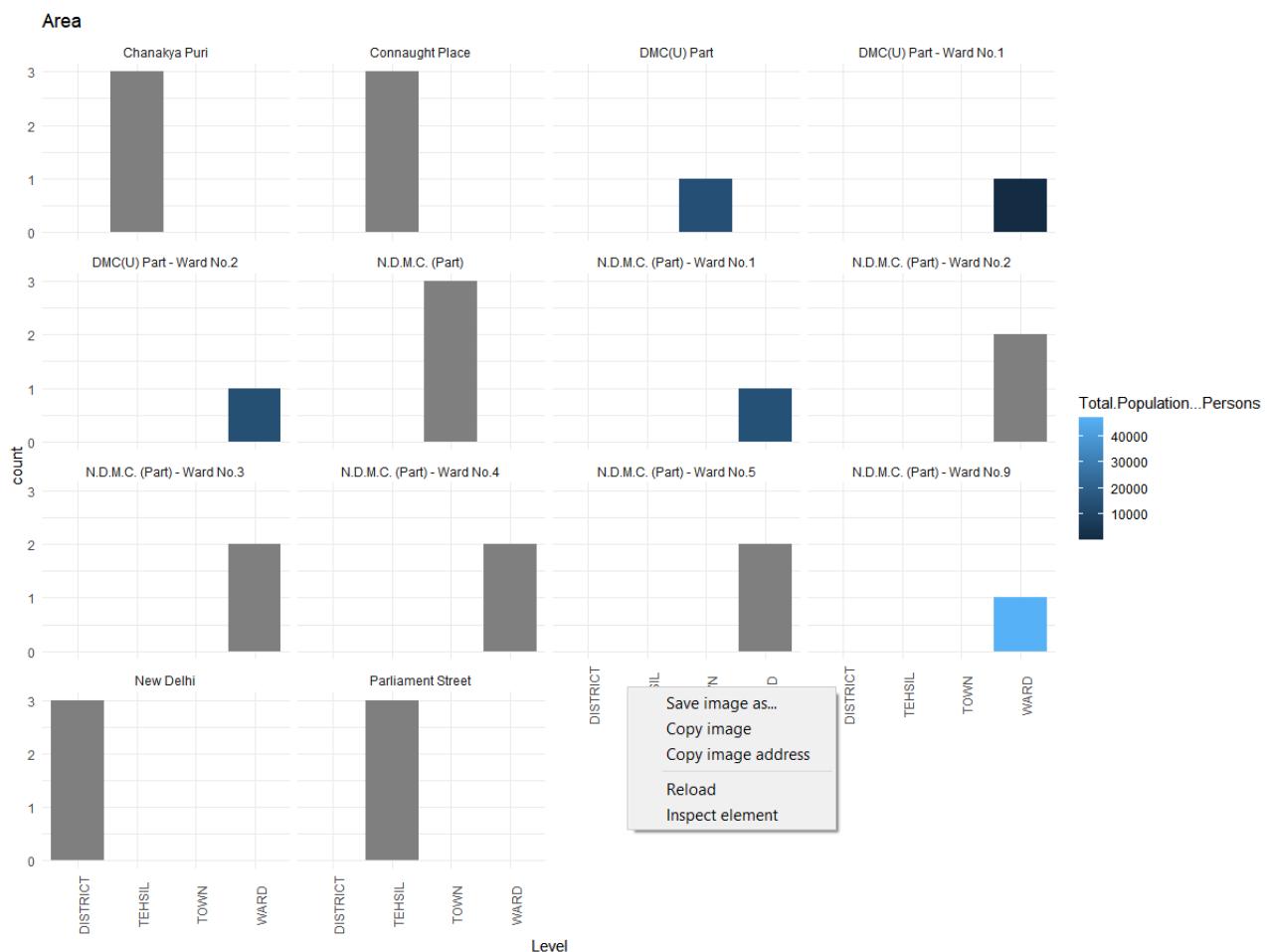
---

```
#density
ggplot(mydata) +
  aes(x = Name) +
  geom_density(adjust = 1L, fill = "#ef562d") +
  labs(x = "population per level", y = "Density", title = "Density Curve") +
  theme(axis.text.x = element_text(angle = 90)) +
  facet_wrap(vars(Level))
```



# 8- AREA PLOT OF LEVEL-WISE POPULATION IN DELHI

```
#area plot
ggplot(mydata) +
  aes(x = Level, fill = Total.Population...Persons) +
  geom_bar() +
  labs(title = "Area") +
  scale_fill_gradient() +
  theme_minimal() +
  facet_wrap(vars(Name)) +
  theme(axis.text.x = element_text(angle = 90))
```

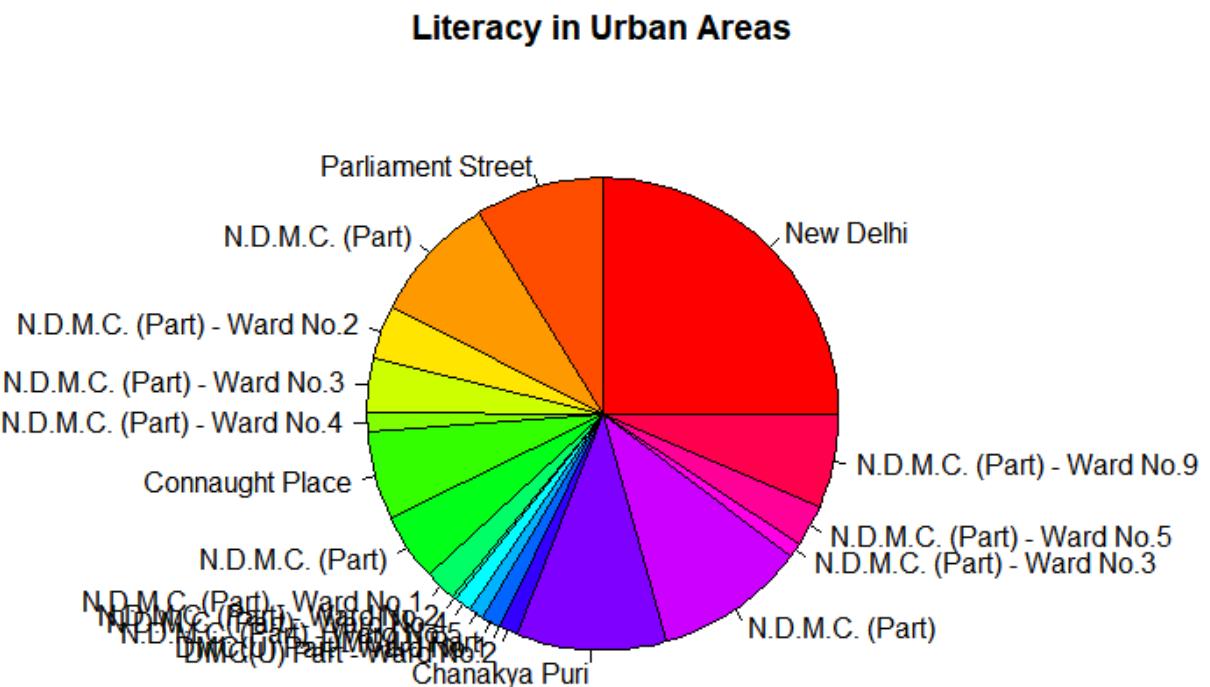


# 9- LITERACY AND ILLITERACY ANALYSIS

---

## PIE-CHART SHOWING LITERACY RATE IN EACH DISTRICT

```
#Literacy rate in each district  
#pie chart  
dataset <- as.data.frame(mydata,stringsasfactors=FALSE)  
# calculating mean of col2 based on col1 group  
  
dataset2 <- (subset(dataset,dataset$Total..Rural..Urban=="Urban"))  
data <- aggregate(dataset2$Illiterate...Persons,  
                   by=list(dataset2$Name), FUN=sum)  
print(data)  
  
pie(data$Illiterate...Persons, data>Name,col =  
rainbow(length(data$Illiterate...Persons)),  
main = "Illiteracy in Urban Areas", cex.lab = 2)
```



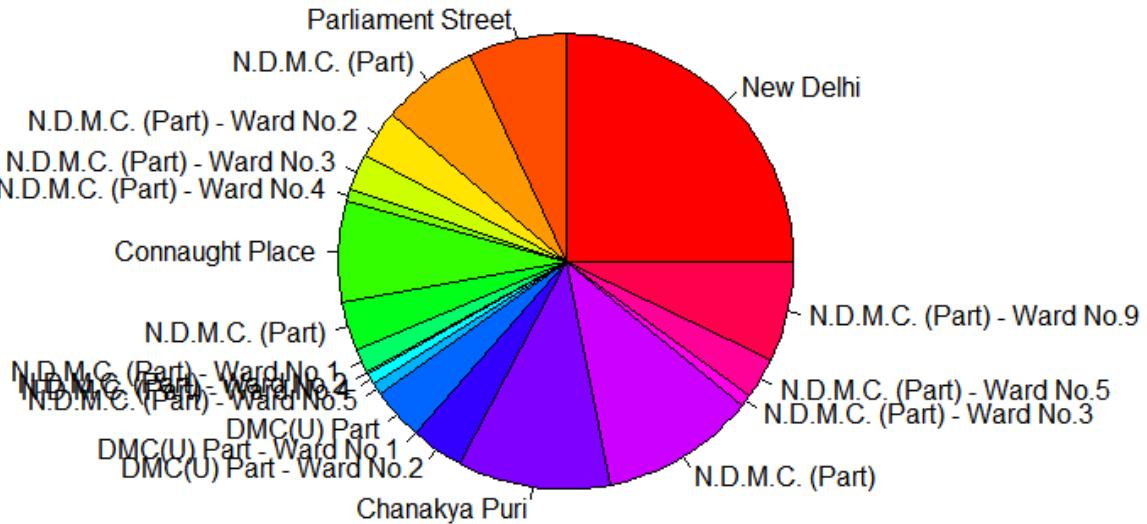
# PIE-CHART SHOWING ILLITERACY RATE IN EACH DISTRICT

```
#Illiteracy rate in each district
dataset <- as.data.frame(mydata,stringsAsFactors=FALSE)
# calculating mean of col2 based on col1 group

dataset2 <- (subset(dataset,dataset$Total..Rural..Urban=="Urban"))
data <- aggregate(dataset2$Literates.Population...Persons,
                  by=list(dataset2$Name), FUN=sum)
print(data)

pie(dataset2$Literates.Population...Persons, dataset2$Name,col =
rainbow(length(dataset2$Literates.Population...Persons)),
main = "Literacy in Urban Areas")
```

**Illiteracy in Urban Areas**

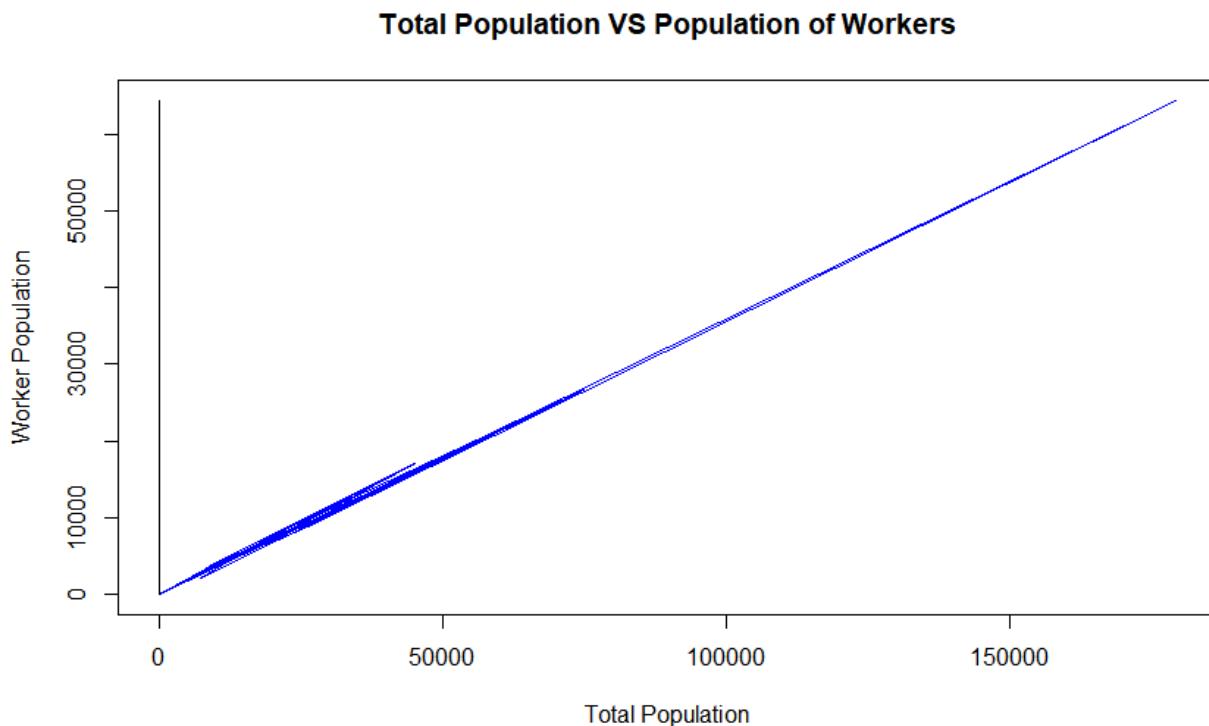


# 10- WORKERS AND NON-WORKERS ANALYSIS

---

## RATIO OF POPULATION OF WORKERS TO TOTAL POPULATION

```
#Total Population VS Population of Workers
plot(mydata$Total.Population...Persons,mydata$Main.Worker.Population...Persons, xlab="Total Population",ylab="Worker Population",main="Total Population VS Population of Workers",type="l",
      col="blue")
lines(mydata$Main.Worker.Population...Persons)
```



# 11- POINT-PLOT FOR TOTAL WORKING POPULATION WITH TOTAL POPULATION OF EACH AREA IN TERMS OF LEVEL

---

```
#point plot
ggplot(mydata) +
  aes(x = Total.Population...Persons, y =
Total.Worker.Population...Persons, colour = Name) +
  geom_point(size = 2L) +
  labs(title = "Point") +
  scale_color_hue() +
  theme_dark() +
  facet_grid(vars(Level), vars())
```

