SMDM Assignment

Harshit

November 10, 2016

# R Assignment - Pune Census Case

Objective of the analysis : The district collector would like to analysis this data with the objective of evaluating the status of the Female population across the district. This statistical analysis should support in taking decisions to run government programs focusing on different areas of Female population such as female literacy, female child welfare, sex ratios, female employment etc.

Below are the inputs to be considered for your analysis:

The CD blocks need to be divided into three categories as below:  
Category A : total population upto 2 lacs  
Category B : total population between 2-5 lacs  
Category C : total population above 5 lacs

The towns need to be divided into three categories as below:  
Category T1 : total population upto 15K  
Category T2 : total population between 15-25K  
Category T3 : total population above 25K

The villages need to be divided into four categories as below:  
Category V1 : total population upto 2500  
Category V2 : total population between 2500-10,000  
Category V3 : total population between 10,000-40,000  
Catgeory V4 : total population above 40,000

Using R execute a statistical analysis as per the above buckets, and help the district collector to understand the following:  
Sex ratio(per 1000)  
Child sex ratio (0-6 age)  
Average literacy rate  
Female literacy rate  
Child proportion(0-6 age)  
Girls proportion(0-6 age)  
Averge percentage of worker population female  
Proportion of Main : Marginal worker population female  
Within female Main worker population the proportion of Main :Main cultivator : Main agri labour: Main houshold industry :Main other  
Average number of child(0-6) per household  
Average number of girl child(0-6) per household  
Average percentage non working population female

Is there a significant difference in the above parameters between the top and the lowest category levels of each of the CD block(A/C), Village(V1/v4) and Towns(T1/T3)?  
Is there a significant difference in the average literacy rate, child sex ratio, Female literacy rate , average female worker population between Urban and Rural ?  
Is there a relationship between the female literacy rate and the average percentage worker population female in the villages?

Overall Rural vs Urban Sex ratio  
Pune district population density: Pune district administers 15,643 square kilometers of areas  
Top 5 highest literacy rate towns, village and district blocks  
Top 5 highest female literacy rate towns, village and district blocks  
Bottom 5 lowest literacy rate towns, village and district blocks  
Bottom 5 lowest female literacy rate towns, village and district blocks  
Lowest sex ratio towns, village and district blocks in each category

## Solution with Output

### Reading CSV file

Census <- read.csv(file.choose())

### Creating subsets for CD Blocks, Towns and Villages

attach(Census)  
CDBlocks <- subset(Census, Level=="CD BLOCK" & Total.Rural.Urban=="Total")  
CDBlocksN <- subset(Census, Level=="CD BLOCK" & Total.Rural.Urban!="Total")  
Town <- subset(Census, Level=="TOWN")  
Village <- subset(Census, Level=="VILLAGE")  
detach(Census)

### Creating Categories in those subsets

attach(CDBlocks)  
CDCat <- ifelse(Total.Population.Person <= 200000, 'A',  
 ifelse(Total.Population.Person > 200000 & Total.Population.Person <= 500000, 'B',  
 ifelse(Total.Population.Person > 500000, 'C','F')))  
  
CDBlockCat <- data.frame(CDBlocks,CDCat)  
detach(CDBlocks)  
  
attach(Town)  
TCat <- ifelse(Total.Population.Person <= 15000, "T1",  
 ifelse(Total.Population.Person > 15000 & Total.Population.Person <= 25000, "T2",  
 ifelse(Total.Population.Person > 25000, "T3", "T9")))  
  
TownCat <- data.frame(Town,TCat)  
detach(Town)  
  
  
attach(Village)  
VCat <- ifelse(Total.Population.Person <= 2500, "V1",  
 ifelse(Total.Population.Person > 2500 & Total.Population.Person <= 10000, "V2",   
 ifelse(Total.Population.Person > 10000 & Total.Population.Person <= 40000, "V3",  
 ifelse(Total.Population.Person > 40000, "V4", "V9"))))  
  
VillageCat <- data.frame(Village,VCat)  
detach(Village)  
  
myvars <- c("CD.Block\_Code", "CDCat")  
CatCD <- CDBlockCat[myvars]  
CDBlockCat2 <- merge(CDBlocksN, CatCD, by="CD.Block\_Code")

### Calculating Ratios

#### CD Block Ratios

library("plyr", lib.loc="~/R/win-library/3.3")  
  
CDBlockRatios <- ddply(CDBlockCat2,"CDCat",summarize,SexRatio=((sum(Total.Population.Female)/sum(Total.Population.Male))\*1000),  
 ChildSexRatio=((sum(Population.in.the.age.group.0.6.Female)/sum(Population.in.the.age.group.0.6.Male))\*1000),  
 AvgLiteracyRate=((mean(Literates.Population.Person)/mean(Total.Population.Person - Population.in.the.age.group.0.6.Person))\*100),  
 FemaleLiteracyRate=((sum(Literates.Population.Female)/sum(Total.Population.Female - Population.in.the.age.group.0.6.Female))\*100),  
 ChildProportion=(sum(Population.in.the.age.group.0.6.Person)/sum(Total.Population.Person)),  
 GirlsProportion=(sum(Population.in.the.age.group.0.6.Female)/sum(Population.in.the.age.group.0.6.Person)),  
 AvgPercentFemaleWorker=((mean(Total.Worker.Population.Female)/mean(Total.Worker.Population.Person))\*100),  
 MainMarginalWorkerFemale=sum(Main.Working.Population.Female)/sum(Marginal.Worker.Population.Female),  
 CultivatorToWorker=(sum(Main.Cultivator.Population.Female)/sum(Main.Working.Population.Female)),  
 AgriLabourToWorker=(sum(Main.Agricultural.Labourers.Population.Female)/sum(Main.Working.Population.Female)),  
 HouseholdToWorker=(sum(Main.Household.Industries.Population.Female)/sum(Main.Working.Population.Female)),  
 OtherToWorker=(sum(Main.Other.Workers.Population.Female)/sum(Main.Working.Population.Female)),  
 AvgChildPerHousehold=(mean(Population.in.the.age.group.0.6.Person)/mean(No.of.Households)),  
 AvgGirlChildPerHousehold=(mean(Population.in.the.age.group.0.6.Female)/mean(No.of.Households)),  
 AvgNonWorkingPopFemale=((mean(Non.Working.Population.Female)/mean(Non.Working.Population.Person))\*100))  
  
#CD Block Ratios  
CDBlockRatios

## CDCat SexRatio ChildSexRatio AvgLiteracyRate FemaleLiteracyRate  
## 1 A 943.6923 888.9674 80.03457 71.61090  
## 2 B 932.0107 868.5137 80.91965 73.26415  
## 3 C 894.3313 857.5973 85.19992 79.26095  
## ChildProportion GirlsProportion AvgPercentFemaleWorker  
## 1 0.1177750 0.4706103 39.46422  
## 2 0.1184131 0.4648153 40.22818  
## 3 0.1347173 0.4616702 27.35288  
## MainMarginalWorkerFemale CultivatorToWorker AgriLabourToWorker  
## 1 4.661909 0.6037235 0.1928253  
## 2 8.191461 0.5445111 0.3046652  
## 3 5.873065 0.3051085 0.2039359  
## HouseholdToWorker OtherToWorker AvgChildPerHousehold  
## 1 0.02691742 0.1765337 0.5429896  
## 2 0.01848877 0.1323349 0.5625962  
## 3 0.05239521 0.4385604 0.5975256  
## AvgGirlChildPerHousehold AvgNonWorkingPopFemale  
## 1 0.2555365 57.84109  
## 2 0.2615033 56.34638  
## 3 0.2758597 61.37866

#### Town Ratios

TownRatios <- ddply(TownCat,"TCat",summarize,SexRatio=((sum(Total.Population.Female)/sum(Total.Population.Male))\*1000),  
 ChildSexRatio=((sum(Population.in.the.age.group.0.6.Female)/sum(Population.in.the.age.group.0.6.Male))\*1000),  
 AvgLiteracyRate=((mean(Literates.Population.Person)/mean(Total.Population.Person - Population.in.the.age.group.0.6.Person))\*100),  
 FemaleLiteracyRate=((sum(Literates.Population.Female)/sum(Total.Population.Female - Population.in.the.age.group.0.6.Female))\*100),  
 ChildProportion=(sum(Population.in.the.age.group.0.6.Person)/sum(Total.Population.Person)),  
 GirlsProportion=(sum(Population.in.the.age.group.0.6.Female)/sum(Population.in.the.age.group.0.6.Person)),  
 AvgPercentFemaleWorker=((mean(Total.Worker.Population.Female)/mean(Total.Worker.Population.Person))\*100),  
 MainMarginalWorkerFemale=sum(Main.Working.Population.Female)/sum(Marginal.Worker.Population.Female),  
 CultivatorToWorker=(sum(Main.Cultivator.Population.Female)/sum(Main.Working.Population.Female)),  
 AgriLabourToWorker=(sum(Main.Agricultural.Labourers.Population.Female)/sum(Main.Working.Population.Female)),  
 HouseholdToWorker=(sum(Main.Household.Industries.Population.Female)/sum(Main.Working.Population.Female)),  
 OtherToWorker=(sum(Main.Other.Workers.Population.Female)/sum(Main.Working.Population.Female)),  
 AvgChildPerHousehold=(mean(Population.in.the.age.group.0.6.Person)/mean(No.of.Households)),  
 AvgGirlChildPerHousehold=(mean(Population.in.the.age.group.0.6.Female)/mean(No.of.Households)),  
 AvgNonWorkingPopFemale=((mean(Non.Working.Population.Female)/mean(Non.Working.Population.Person))\*100))  
  
#Town Ratios  
TownRatios

## TCat SexRatio ChildSexRatio AvgLiteracyRate FemaleLiteracyRate  
## 1 T1 850.5653 847.3477 87.20540 82.10408  
## 2 T2 885.0100 888.4236 89.37312 84.97646  
## 3 T3 886.8331 854.7531 87.64341 83.87735  
## ChildProportion GirlsProportion AvgPercentFemaleWorker  
## 1 0.1436578 0.4586834 21.17377  
## 2 0.1172679 0.4704578 25.01466  
## 3 0.1400670 0.4608447 22.66477  
## MainMarginalWorkerFemale CultivatorToWorker AgriLabourToWorker  
## 1 5.214958 0.1011492 0.07455599  
## 2 6.172783 0.1245975 0.08645033  
## 3 6.079774 0.1039099 0.06746190  
## HouseholdToWorker OtherToWorker AvgChildPerHousehold  
## 1 0.07693038 0.7473644 0.5867985  
## 2 0.05697300 0.7319792 0.5529958  
## 3 0.07037773 0.7582505 0.5848859  
## AvgGirlChildPerHousehold AvgNonWorkingPopFemale  
## 1 0.2691547 62.77074  
## 2 0.2601612 58.67564  
## 3 0.2695416 62.47355

#### Village Ratios

VillageRatios <- ddply(VillageCat,"VCat",summarize,SexRatio=((sum(Total.Population.Female)/sum(Total.Population.Male))\*1000),  
 ChildSexRatio=((sum(Population.in.the.age.group.0.6.Female)/sum(Population.in.the.age.group.0.6.Male))\*1000),  
 AvgLiteracyRate=((mean(Literates.Population.Person)/mean(Total.Population.Person - Population.in.the.age.group.0.6.Person))\*100),  
 FemaleLiteracyRate=((sum(Literates.Population.Female)/sum(Total.Population.Female - Population.in.the.age.group.0.6.Female))\*100),  
 ChildProportion=(sum(Population.in.the.age.group.0.6.Person)/sum(Total.Population.Person)),  
 GirlsProportion=(sum(Population.in.the.age.group.0.6.Female)/sum(Population.in.the.age.group.0.6.Person)),  
 AvgPercentFemaleWorker=((mean(Total.Worker.Population.Female)/mean(Total.Worker.Population.Person))\*100),  
 MainMarginalWorkerFemale=sum(Main.Working.Population.Female)/sum(Marginal.Worker.Population.Female),  
 CultivatorToWorker=(sum(Main.Cultivator.Population.Female)/sum(Main.Working.Population.Female)),  
 AgriLabourToWorker=(sum(Main.Agricultural.Labourers.Population.Female)/sum(Main.Working.Population.Female)),  
 HouseholdToWorker=(sum(Main.Household.Industries.Population.Female)/sum(Main.Working.Population.Female)),  
 OtherToWorker=(sum(Main.Other.Workers.Population.Female)/sum(Main.Working.Population.Female)),  
 AvgChildPerHousehold=(mean(Population.in.the.age.group.0.6.Person)/mean(No.of.Households)),  
 AvgGirlChildPerHousehold=(mean(Population.in.the.age.group.0.6.Female)/mean(No.of.Households)),  
 AvgNonWorkingPopFemale=((mean(Non.Working.Population.Female)/mean(Non.Working.Population.Person))\*100))  
  
#Village Ratios  
VillageRatios

## VCat SexRatio ChildSexRatio AvgLiteracyRate FemaleLiteracyRate  
## 1 V1 954.5543 889.5897 78.37244 69.40420  
## 2 V2 926.9032 858.6706 81.41359 74.04796  
## 3 V3 897.9564 862.6513 85.41298 79.71637  
## 4 V4 901.6667 854.2000 88.68483 84.14535  
## ChildProportion GirlsProportion AvgPercentFemaleWorker  
## 1 0.1148131 0.4707846 43.62064  
## 2 0.1193783 0.4619811 39.14817  
## 3 0.1298270 0.4631309 28.32958  
## 4 0.1403379 0.4606839 20.96456  
## MainMarginalWorkerFemale CultivatorToWorker AgriLabourToWorker  
## 1 6.032640 0.6345257 0.2593510  
## 2 9.124164 0.5166513 0.3139195  
## 3 6.764486 0.2792433 0.2779616  
## 4 4.646804 0.1492189 0.2056423  
## HouseholdToWorker OtherToWorker AvgChildPerHousehold  
## 1 0.01388639 0.09223691 0.5584908  
## 2 0.02488085 0.14454834 0.5635821  
## 3 0.03971576 0.40307927 0.5751619  
## 4 0.04639776 0.59874097 0.5944854  
## AvgGirlChildPerHousehold AvgNonWorkingPopFemale  
## 1 0.2629289 54.84750  
## 2 0.2603643 56.91948  
## 3 0.2663752 61.08833  
## 4 0.2738698 63.37468

#### Rural vs Urban Ratios

RuralUrbanRatios <- ddply(CDBlockCat2,"Total.Rural.Urban",summarize,SexRatio=((sum(Total.Population.Female)/sum(Total.Population.Male))\*1000),  
 ChildSexRatio=((sum(Population.in.the.age.group.0.6.Female)/sum(Population.in.the.age.group.0.6.Male))\*1000),  
 AvgLiteracyRate=((mean(Literates.Population.Person)/mean(Total.Population.Person))\*100),  
 FemaleLiteracyRate=((sum(Literates.Population.Female)/sum(Total.Population.Female))\*100),  
 ChildProportion=(sum(Population.in.the.age.group.0.6.Person)/sum(Total.Population.Person)),  
 GirlsProportion=(sum(Population.in.the.age.group.0.6.Female)/sum(Population.in.the.age.group.0.6.Person)),  
 AvgPercentFemaleWorker=((mean(Total.Worker.Population.Female)/mean(Total.Worker.Population.Person))\*100),  
 MainMarginalWorkerFemale=sum(Main.Working.Population.Female)/sum(Marginal.Worker.Population.Female),  
 CultivatorToWorker=(sum(Main.Cultivator.Population.Female)/sum(Main.Working.Population.Female)),  
 AgriLabourToWorker=(sum(Main.Agricultural.Labourers.Population.Female)/sum(Main.Working.Population.Female)),  
 HouseholdToWorker=(sum(Main.Household.Industries.Population.Female)/sum(Main.Working.Population.Female)),  
 OtherToWorker=(sum(Main.Other.Workers.Population.Female)/sum(Main.Working.Population.Female)),  
 AvgChildPerHousehold=(mean(Population.in.the.age.group.0.6.Person)/mean(No.of.Households)),  
 AvgGirlChildPerHousehold=(mean(Population.in.the.age.group.0.6.Female)/mean(No.of.Households)),  
 AvgNonWorkingPopFemale=((mean(Non.Working.Population.Female)/mean(Non.Working.Population.Person))\*100))  
  
#Rural vs Urban Ratios  
RuralUrbanRatios

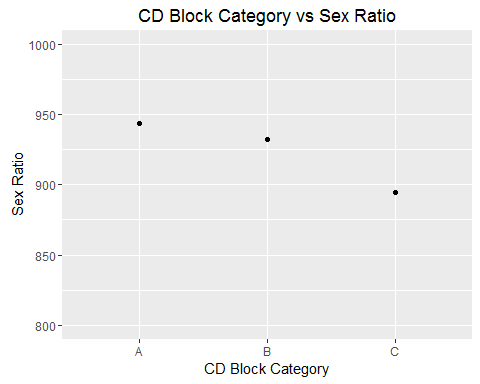
## Total.Rural.Urban SexRatio ChildSexRatio AvgLiteracyRate  
## 1 Rural 932.4097 870.8002 71.29040  
## 2 Urban 868.6197 856.0144 75.66257  
## FemaleLiteracyRate ChildProportion GirlsProportion  
## 1 64.79829 0.1196506 0.4654694  
## 2 71.85695 0.1377330 0.4612111  
## AvgPercentFemaleWorker MainMarginalWorkerFemale CultivatorToWorker  
## 1 39.31287 7.153319 0.5449746  
## 2 22.28549 5.649208 0.1063724  
## AgriLabourToWorker HouseholdToWorker OtherToWorker AvgChildPerHousehold  
## 1 0.28465945 0.02143805 0.1489279 0.5642536  
## 2 0.07430691 0.07105061 0.7482701 0.5807376  
## AvgGirlChildPerHousehold AvgNonWorkingPopFemale  
## 1 0.2626428 57.08768  
## 2 0.2678426 61.88888

### *Question : Is there a significant difference in the above parameters between the top and the lowest category levels of each of the CD block(A/C), Village(V1/v4) and Towns(T1/T3)?*

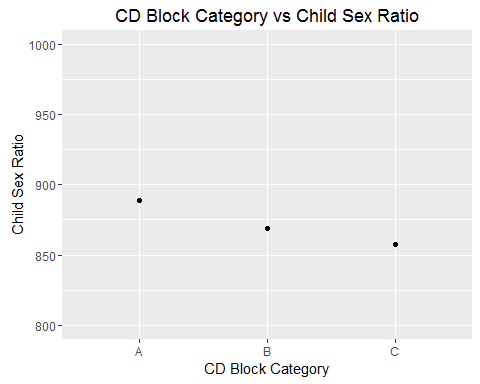
**Check the graphs below for the difference in the above parameters between top and lowest category levels of CD Blocks**

## The following object is masked \_by\_ .GlobalEnv:  
##   
## CDCat

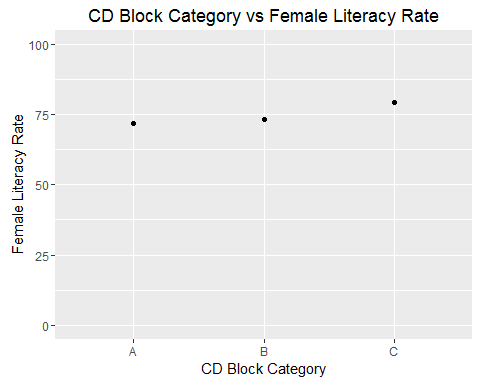
qplot(CDCat, SexRatio, data=CDBlockRatios, xlab="CD Block Category", ylab="Sex Ratio", main="CD Block Category vs Sex Ratio", ylim=c(800,1000))



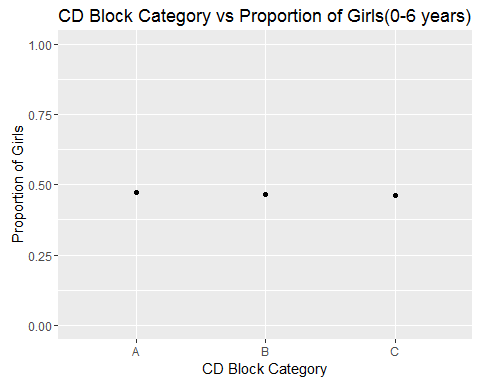
qplot(CDCat, ChildSexRatio, data=CDBlockRatios, xlab="CD Block Category", ylab="Child Sex Ratio", main="CD Block Category vs Child Sex Ratio", ylim=c(800,1000))



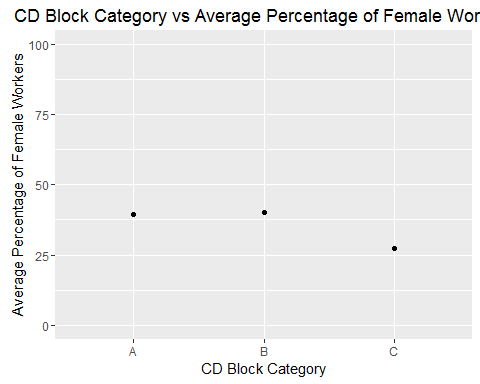
qplot(CDCat, FemaleLiteracyRate, data=CDBlockRatios, xlab="CD Block Category", ylab="Female Literacy Rate", main="CD Block Category vs Female Literacy Rate", ylim=c(0,100))



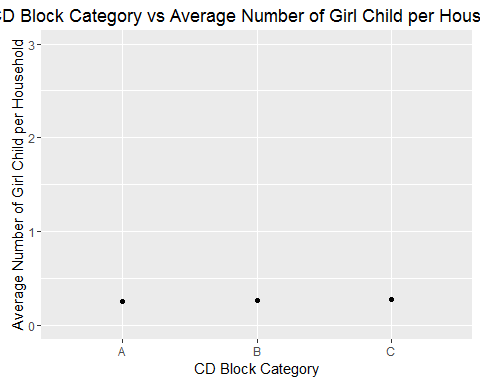
qplot(CDCat, GirlsProportion, data=CDBlockRatios, xlab="CD Block Category", ylab="Proportion of Girls", main="CD Block Category vs Proportion of Girls(0-6 years)", ylim=c(0,1))



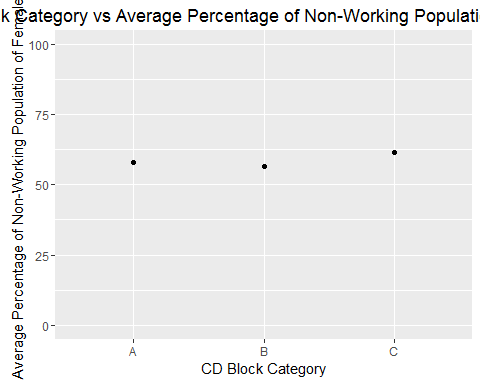
qplot(CDCat, AvgPercentFemaleWorker, data=CDBlockRatios, xlab="CD Block Category", ylab="Average Percentage of Female Workers", main="CD Block Category vs Average Percentage of Female Workers", ylim=c(0,100))



qplot(CDCat, AvgGirlChildPerHousehold, data=CDBlockRatios, xlab="CD Block Category", ylab="Average Number of Girl Child per Household", main="CD Block Category vs Average Number of Girl Child per Household", ylim=c(0,3))



qplot(CDCat, AvgNonWorkingPopFemale, data=CDBlockRatios, xlab="CD Block Category", ylab="Average Percentage of Non-Working Population of Females", main="CD Block Category vs Average Percentage of Non-Working Population of Females", ylim=c(0,100))

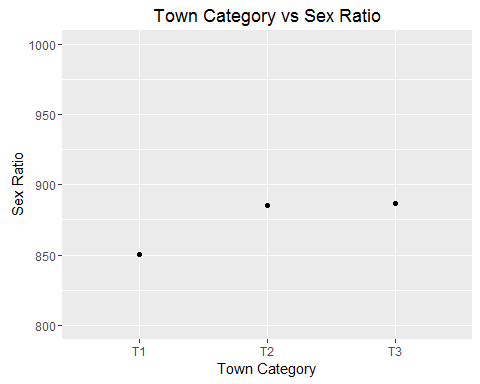


**There is significant difference in Sex Ratio and Child Sex Ratio between CD Block category A and C. The other ratios are similar.**

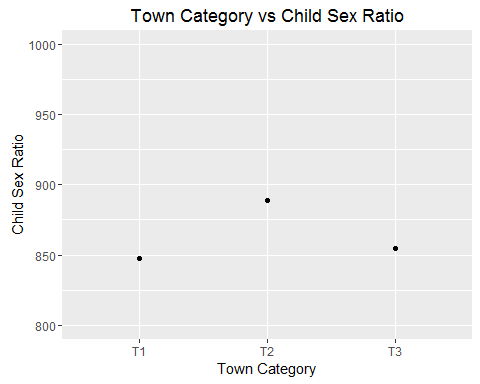
**Check the graphs below for the difference in the above parameters between top and lowest category levels of Towns**

## The following object is masked \_by\_ .GlobalEnv:  
##   
## TCat

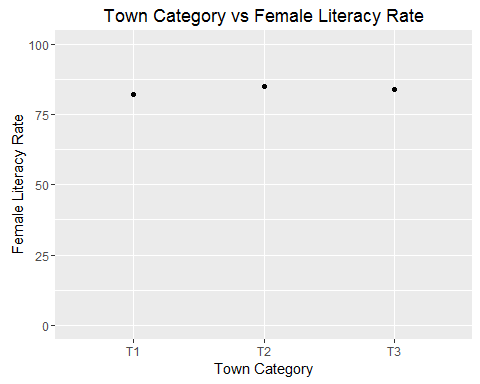
qplot(TCat, SexRatio, data=TownRatios, xlab="Town Category", ylab="Sex Ratio", main="Town Category vs Sex Ratio", ylim=c(800,1000))



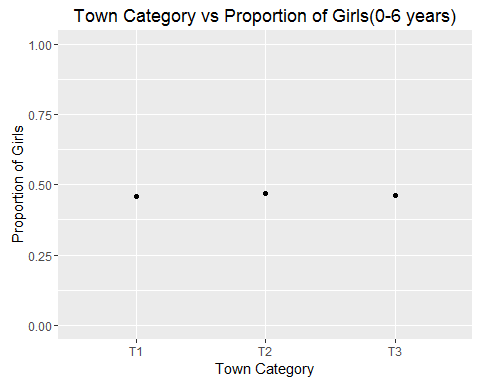
qplot(TCat, ChildSexRatio, data=TownRatios, xlab="Town Category", ylab="Child Sex Ratio", main="Town Category vs Child Sex Ratio", ylim=c(800,1000))



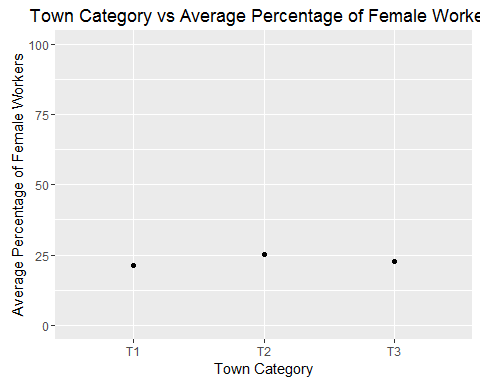
qplot(TCat, FemaleLiteracyRate, data=TownRatios, xlab="Town Category", ylab="Female Literacy Rate", main="Town Category vs Female Literacy Rate", ylim=c(0,100))



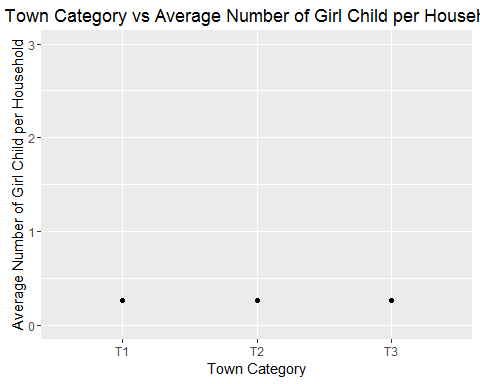
qplot(TCat, GirlsProportion, data=TownRatios, xlab="Town Category", ylab="Proportion of Girls", main="Town Category vs Proportion of Girls(0-6 years)", ylim=c(0,1))



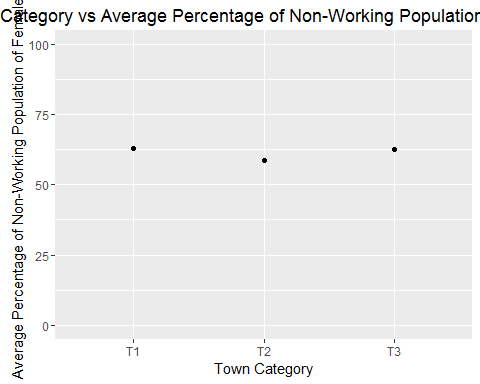
qplot(TCat, AvgPercentFemaleWorker, data=TownRatios, xlab="Town Category", ylab="Average Percentage of Female Workers", main="Town Category vs Average Percentage of Female Workers", ylim=c(0,100))



qplot(TCat, AvgGirlChildPerHousehold, data=TownRatios, xlab="Town Category", ylab="Average Number of Girl Child per Household", main="Town Category vs Average Number of Girl Child per Household", ylim=c(0,3))



qplot(TCat, AvgNonWorkingPopFemale, data=TownRatios, xlab="Town Category", ylab="Average Percentage of Non-Working Population of Females", main="Town Category vs Average Percentage of Non-Working Population of Females", ylim=c(0,100))

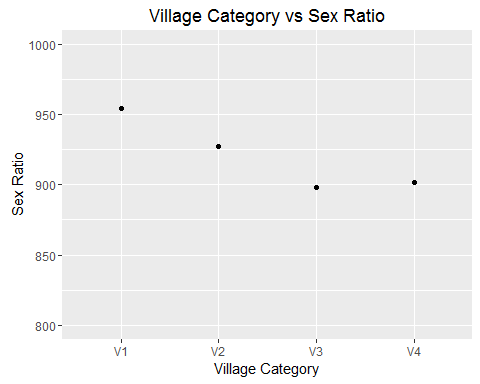


**There is significant difference in Sex Ratio between Town category T1 and T3. The rest of the ratios do not have much variation between the two.**

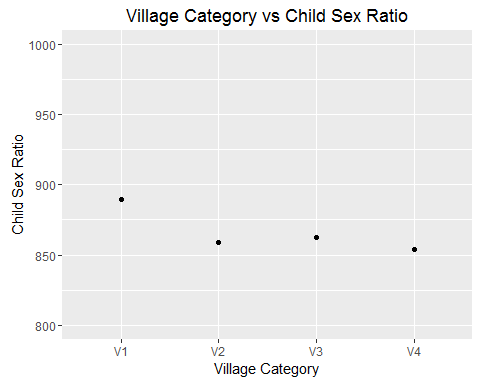
**Check the graphs below for the difference in the above parameters between top and lowest category levels of Villages**

## The following object is masked \_by\_ .GlobalEnv:  
##   
## VCat

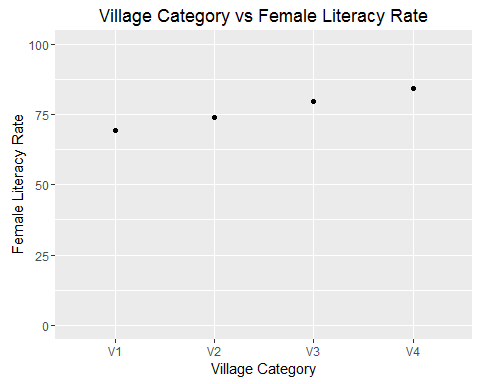
qplot(VCat, SexRatio, data=VillageRatios, xlab="Village Category", ylab="Sex Ratio", main="Village Category vs Sex Ratio", ylim=c(800,1000))



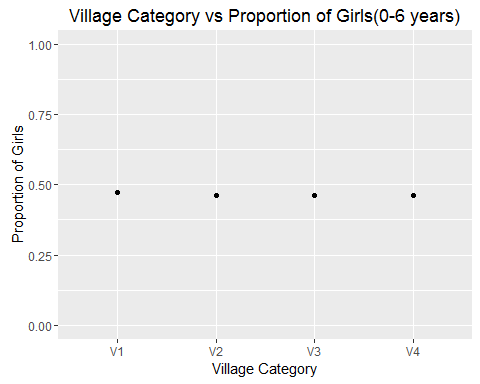
qplot(VCat, ChildSexRatio, data=VillageRatios, xlab="Village Category", ylab="Child Sex Ratio", main="Village Category vs Child Sex Ratio", ylim=c(800,1000))



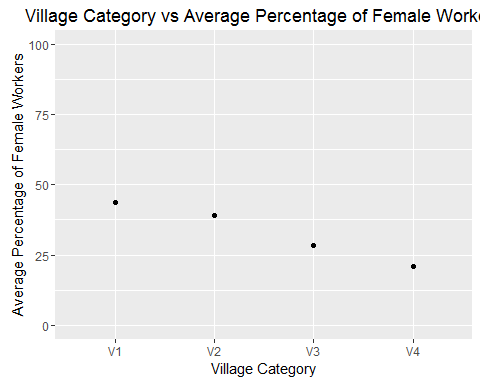
qplot(VCat, FemaleLiteracyRate, data=VillageRatios, xlab="Village Category", ylab="Female Literacy Rate", main="Village Category vs Female Literacy Rate", ylim=c(0,100))



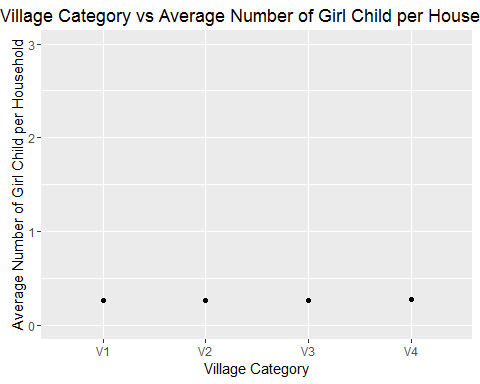
qplot(VCat, GirlsProportion, data=VillageRatios, xlab="Village Category", ylab="Proportion of Girls", main="Village Category vs Proportion of Girls(0-6 years)", ylim=c(0,1))



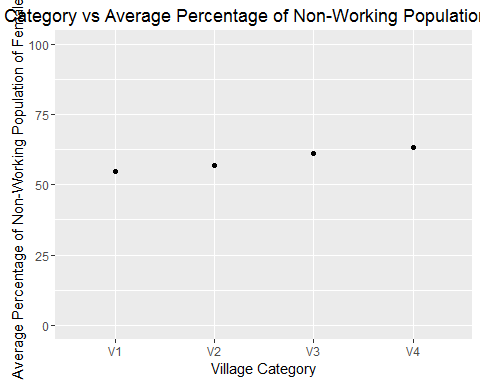
qplot(VCat, AvgPercentFemaleWorker, data=VillageRatios, xlab="Village Category", ylab="Average Percentage of Female Workers", main="Village Category vs Average Percentage of Female Workers", ylim=c(0,100))



qplot(VCat, AvgGirlChildPerHousehold, data=VillageRatios, xlab="Village Category", ylab="Average Number of Girl Child per Household", main="Village Category vs Average Number of Girl Child per Household", ylim=c(0,3))



qplot(VCat, AvgNonWorkingPopFemale, data=VillageRatios, xlab="Village Category", ylab="Average Percentage of Non-Working Population of Females", main="Village Category vs Average Percentage of Non-Working Population of Females", ylim=c(0,100))

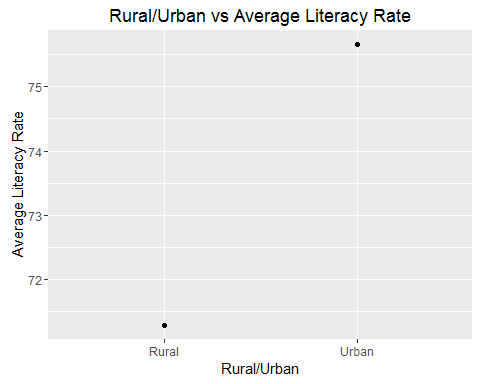


**There is significant difference in Sex Ratio, Child Sex Ratio, Female Literacy Rate and Average Percenatge of Female Workers between Villages of category V1 and V4. There is not much defference in the other ratios.**

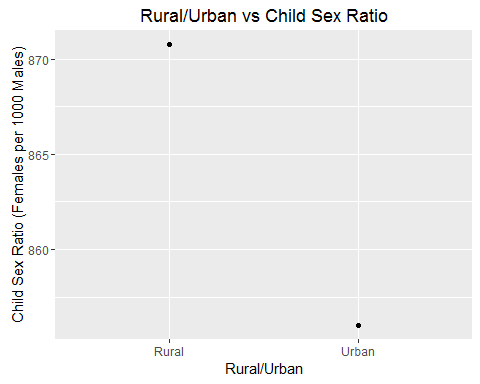
### *Question : Is there a significant difference in the average literacy rate, child sex ratio, Female literacy rate , average female worker population between Urban and Rural?*

**Check the graphs below for the differences between Rural and Urban**

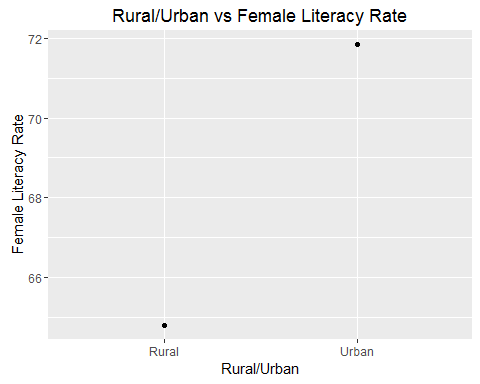
qplot(Total.Rural.Urban, AvgLiteracyRate, data=RuralUrbanRatios, xlab="Rural/Urban", ylab="Average Literacy Rate", main="Rural/Urban vs Average Literacy Rate")



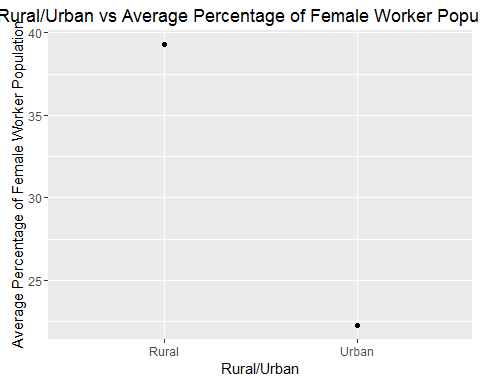
qplot(Total.Rural.Urban, ChildSexRatio, data=RuralUrbanRatios, xlab="Rural/Urban", ylab="Child Sex Ratio (Females per 1000 Males)", main="Rural/Urban vs Child Sex Ratio")



qplot(Total.Rural.Urban, FemaleLiteracyRate, data=RuralUrbanRatios, xlab="Rural/Urban", ylab="Female Literacy Rate", main="Rural/Urban vs Female Literacy Rate")



qplot(Total.Rural.Urban, AvgPercentFemaleWorker, data=RuralUrbanRatios, xlab="Rural/Urban", ylab="Average Percentage of Female Worker Population", main="Rural/Urban vs Average Percentage of Female Worker Population")



**As we can see from the graphs above, there is very significant difference between rural and urban for the mentioned ratios.**

### *Question : Is there a relationship between the female literacy rate and the average percentage worker population female in the villages?*

VFemaleLiteracyRate <- VillageRatios$FemaleLiteracyRate  
VAvgPercentWorkingPopFemale <- VillageRatios$AvgPercentFemaleWorker  
cor(VFemaleLiteracyRate, VAvgPercentWorkingPopFemale)

## [1] -0.99281

**As we can see above there is a strong relationship between the two as there is a nearly perfect, linear negative correlation.**

#### Rural vs Urban Sex Ratio

RuralUrbanRatios[1:2]

## Total.Rural.Urban SexRatio  
## 1 Rural 932.4097  
## 2 Urban 868.6197

#### Pune Population Density

PunePopulationDensity <- sum(CDBlockCat2$Total.Population.Person)/15643  
PunePopulationDensity

## [1] 254.3284

#### Top 5 Literacy Rates

**Top 5 CD Blocks based on Literacy Rates**

#Top 5 CD Block Literacy Rates  
LiteracyRate <- (CDBlockCat$Literates.Population.Person/CDBlockCat$Total.Population.Person)\*100  
CDBlockCat <- data.frame(CDBlockCat, LiteracyRate)  
Top5LiteracyRateCDBlock <- head(arrange(CDBlockCat,desc(LiteracyRate)), n = 5)  
Top5LiteracyRateCDBlock[c(3,10,97,98)]

## CD.Block\_Code Name CDCat LiteracyRate  
## 1 245 Pune A 75.75505  
## 2 238 Junnar B 74.38120  
## 3 239 Ambegaon B 74.36857  
## 4 244 Haveli C 73.72202  
## 5 247 Purandhar A 72.64463

**Top 5 Towns based on Literacy Rates**

#Top 5 Town Literacy Rates  
LiteracyRate <- (TownCat$Literates.Population.Person/TownCat$Total.Population.Person)\*100  
TownCat <- data.frame(TownCat, LiteracyRate)  
Top5LiteracyRateTown <- head(arrange(TownCat,desc(LiteracyRate)), n = 5)  
Top5LiteracyRateTown[c(3,10,97,98)]

## CD.Block\_Code Name TCat LiteracyRate  
## 1 241 Rajgurunagar (Khed) (CT) T3 80.27519  
## 2 250 Jalochi (CT) T2 79.24636  
## 3 244 Dehu (CT) T1 79.09669  
## 4 239 Manchar (CT) T2 78.80907  
## 5 242 Kusgaon Bk. (CT) T2 78.55496

**Top 5 Villages based on Literacy Rates**

#Top 5 Village Literacy Rates  
LiteracyRate <- (VillageCat$Literates.Population.Person/VillageCat$Total.Population.Person)\*100  
VillageCat <- data.frame(VillageCat, LiteracyRate)  
Top5LiteracyRateVillage <- head(arrange(VillageCat,desc(LiteracyRate)), n = 5)  
Top5LiteracyRateVillage[c(3,10,97,98)]

## CD.Block\_Code Name VCat LiteracyRate  
## 1 249 Bhabavadi V1 90.83821  
## 2 247 Ambodi V1 87.07692  
## 3 239 Gadewadi V1 87.04512  
## 4 239 Terungan V1 86.18267  
## 5 243 Padalghar V1 85.71429

#### Top 5 Female Literacy Rates

**Top 5 CD Blocks based on Female Literacy Rates**

#Top 5 CD Block Female Literacy Rates  
LiteracyRateFemale <- (CDBlockCat$Literates.Population.Female/CDBlockCat$Total.Population.Female)\*100  
CDBlockCat <- data.frame(CDBlockCat, LiteracyRateFemale)  
Top5FemaleLiteracyRateCDBlock <- head(arrange(CDBlockCat,desc(LiteracyRateFemale)), n = 5)  
Top5FemaleLiteracyRateCDBlock[c(3,10,97,99)]

## CD.Block\_Code Name CDCat LiteracyRateFemale  
## 1 245 Pune A 73.19859  
## 2 244 Haveli C 68.81923  
## 3 238 Junnar B 68.19858  
## 4 239 Ambegaon B 67.77534  
## 5 250 Baramati B 66.03552

**Top 5 Towns based on Female Literacy Rates**

#Top 5 Town Female Literacy Rates  
LiteracyRateFemale <- (TownCat$Literates.Population.Female/TownCat$Total.Population.Female)\*100  
TownCat <- data.frame(TownCat, LiteracyRateFemale)  
Top5FemaleLiteracyRateTown <- head(arrange(TownCat,desc(LiteracyRateFemale)), n = 5)  
Top5FemaleLiteracyRateTown[c(3,10,97,99)]

## CD.Block\_Code Name TCat LiteracyRateFemale  
## 1 241 Rajgurunagar (Khed) (CT) T3 78.54168  
## 2 250 Jalochi (CT) T2 77.26737  
## 3 244 Dehu (CT) T1 75.70220  
## 4 239 Manchar (CT) T2 75.59840  
## 5 242 Wadagaon (CT) T1 72.21905

**Top 5 Villages based on Female Literacy Rates**

#Top 5 Village Female Literacy Rates  
LiteracyRateFemale <- (VillageCat$Literates.Population.Female/VillageCat$Total.Population.Female)\*100  
VillageCat <- data.frame(VillageCat, LiteracyRateFemale)  
Top5FemaleLiteracyRateVillage <- head(arrange(VillageCat,desc(LiteracyRateFemale)), n = 5)  
Top5FemaleLiteracyRateVillage[c(3,10,97,99)]

## CD.Block\_Code Name VCat LiteracyRateFemale  
## 1 249 Bhabavadi V1 92.71255  
## 2 247 Ambodi V1 87.50000  
## 3 239 Gadewadi V1 86.98061  
## 4 239 Terungan V1 82.89474  
## 5 249 Panjalwadi V1 80.96447

#### Bottom 5 Literacy Rates

**Bottom 5 CD Blocks based on Literacy Rates**

#Bottom 5 CD Block Literacy Rates  
Bottom5LiteracyRateCDBlock <- head(arrange(CDBlockCat,LiteracyRate), n = 5)  
Bottom5LiteracyRateCDBlock[c(3,10,97,98)]

## CD.Block\_Code Name CDCat LiteracyRate  
## 1 248 Velhe A 67.30134  
## 2 243 Mulshi A 67.78300  
## 3 246 Daund B 68.19143  
## 4 242 Mawal B 68.51450  
## 5 241 Khed B 70.08810

**Bottom Towns based on Literacy Rates**

#Bottom 5 Town Literacy Rates  
Bottom5LiteracyRateTown <- head(arrange(TownCat,LiteracyRate), n = 5)  
Bottom5LiteracyRateTown[c(3,10,97,98)]

## CD.Block\_Code Name TCat LiteracyRate  
## 1 244 Yewalewadi (CT) T1 66.51919  
## 2 243 Hinjavadi (CT) T1 71.30640  
## 3 244 Wagholi (CT) T3 72.63957  
## 4 243 Pirangut (CT) T1 73.53605  
## 5 241 Kharabwadi (CT) T1 73.68478

**Bottom 5 Villages based on Literacy Rates**

#Bottom 5 Village Literacy Rates  
Bottom5LiteracyRateVillage <- head(arrange(VillageCat,LiteracyRate), n = 5)  
Bottom5LiteracyRateVillage[c(3,10,97,98)]

## CD.Block\_Code Name VCat LiteracyRate  
## 1 241 Vanjul Vihire V1 26.66667  
## 2 248 Mohari V1 30.76923  
## 3 243 Mugaon V1 30.84112  
## 4 249 Bope V1 32.19512  
## 5 248 Gondekhal V1 32.87671

#### Bottom 5 Female Literacy Rates

**Bottom 5 CD Blocks based on Female Literacy Rates**

#Bottom 5 CD Block Female Literacy Rates  
Bottom5FemaleLiteracyRateCDBlock <- head(arrange(CDBlockCat,LiteracyRateFemale), n = 5)  
Bottom5FemaleLiteracyRateCDBlock[c(3,10,97,99)]

## CD.Block\_Code Name CDCat LiteracyRateFemale  
## 1 248 Velhe A 58.58878  
## 2 243 Mulshi A 60.66236  
## 3 242 Mawal B 60.77340  
## 4 246 Daund B 62.08782  
## 5 241 Khed B 63.06391

**Bottom 5 CD Blocks based on Female Literacy Rates**

#Bottom 5 Town Female Literacy Rates  
Bottom5FemaleLiteracyRateTown <- head(arrange(TownCat,LiteracyRateFemale), n = 5)  
Bottom5FemaleLiteracyRateTown[c(3,10,97,99)]

## CD.Block\_Code Name TCat LiteracyRateFemale  
## 1 244 Yewalewadi (CT) T1 59.94927  
## 2 243 Hinjavadi (CT) T1 65.40174  
## 3 244 Wagholi (CT) T3 68.25993  
## 4 241 Kharabwadi (CT) T1 69.97840  
## 5 240 Sanaswadi (CT) T1 70.00000

**Bottom 5 CD Blocks based on Female Literacy Rates**

#Bottom 5 Village Female Literacy Rates  
Bottom5FemaleLiteracyRateVillage <- head(arrange(VillageCat,LiteracyRateFemale), n = 5)  
Bottom5FemaleLiteracyRateVillage[c(3,10,97,99)]

## CD.Block\_Code Name VCat LiteracyRateFemale  
## 1 243 Bembatmal V1 23.07692  
## 2 243 Sakhari V1 25.00000  
## 3 248 Mohari V1 25.00000  
## 4 248 Pishawi V1 25.00000  
## 5 243 Nandivali V1 26.49573

#### Category Wise Lowest Sex Ratios

**Category Wise Lowest Sex Ratios in CD Blocks**

#Lowest Sex Ratio CD Blocks  
SexRatio <- (CDBlockCat$Total.Population.Female/CDBlockCat$Total.Population.Male)\*1000  
CDBlockCat <- data.frame(CDBlockCat, SexRatio)  
CDBlockCat1 <- subset(CDBlockCat, CDCat=="A")  
CDBlockCat2 <- subset(CDBlockCat, CDCat=="B")  
CDBlockCat3 <- subset(CDBlockCat, CDCat=="C")  
LowestSexRatioCDBlockA <- head(arrange(CDBlockCat1,SexRatio), n = 1)  
LowestSexRatioCDBlockA[c(3,10,97,100)]

## CD.Block\_Code Name CDCat SexRatio  
## 1 245 Pune A 837.3291

LowestSexRatioCDBlockB <- head(arrange(CDBlockCat2,SexRatio), n = 1)  
LowestSexRatioCDBlockB[c(3,10,97,100)]

## CD.Block\_Code Name CDCat SexRatio  
## 1 242 Mawal B 900.6205

LowestSexRatioCDBlockC <- head(arrange(CDBlockCat3,SexRatio), n = 1)  
LowestSexRatioCDBlockC[c(3,10,97,100)]

## CD.Block\_Code Name CDCat SexRatio  
## 1 244 Haveli C 894.3313

**Category Wise Lowest Sex Ratios in Towns**

#Lowest Sex Ratio Towns  
SexRatio <- (TownCat$Total.Population.Female/TownCat$Total.Population.Male)\*1000  
TownCat <- data.frame(TownCat, SexRatio)  
TownCat1 <- subset(TownCat, TCat=="T1")  
TownCat2 <- subset(TownCat, TCat=="T2")  
TownCat3 <- subset(TownCat, TCat=="T3")  
LowestSexRatioTownT1 <- head(arrange(TownCat1,SexRatio), n = 1)  
LowestSexRatioTownT1[c(3,9,10,97,100)]

## CD.Block\_Code Level Name TCat SexRatio  
## 1 241 TOWN Nanekarwadi (CT) T1 700.3494

LowestSexRatioTownT2 <- head(arrange(TownCat2,SexRatio), n = 1)  
LowestSexRatioTownT2[c(3,9,10,97,100)]

## CD.Block\_Code Level Name TCat SexRatio  
## 1 242 TOWN Kusgaon Bk. (CT) T2 752.3852

LowestSexRatioTownT3 <- head(arrange(TownCat3,SexRatio), n = 1)  
LowestSexRatioTownT3[c(3,9,10,97,100)]

## CD.Block\_Code Level Name TCat SexRatio  
## 1 241 TOWN Chakan (CT) T3 852.4376

**Category Wise Lowest Sex Ratios in Villages**

#Lowest Sex Ratio Villages  
SexRatio <- (VillageCat$Total.Population.Female/VillageCat$Total.Population.Male)\*1000  
VillageCat <- data.frame(VillageCat, SexRatio)  
VillageCat1 <- subset(VillageCat, VCat=="V1")  
VillageCat2 <- subset(VillageCat, VCat=="V2")  
VillageCat3 <- subset(VillageCat, VCat=="V3")  
VillageCat4 <- subset(VillageCat, VCat=="V4")  
LowestSexRatioVillageV1 <- head(arrange(VillageCat1,SexRatio), n = 1)  
LowestSexRatioVillageV1[c(3,9,10,97,100)]

## CD.Block\_Code Level Name VCat SexRatio  
## 1 242 VILLAGE Takave Kh. V1 696.9388

LowestSexRatioVillageV2 <- head(arrange(VillageCat2,SexRatio), n = 1)  
LowestSexRatioVillageV2[c(3,9,10,97,100)]

## CD.Block\_Code Level Name VCat SexRatio  
## 1 241 VILLAGE Mahalunge V2 516.6745

LowestSexRatioVillageV3 <- head(arrange(VillageCat3,SexRatio), n = 1)  
LowestSexRatioVillageV3[c(3,9,10,97,100)]

## CD.Block\_Code Level Name VCat SexRatio  
## 1 240 VILLAGE Ranjangaon Ganpati V3 668.3648

LowestSexRatioVillageV4 <- head(arrange(VillageCat4,SexRatio), n = 1)  
LowestSexRatioVillageV4[c(3,9,10,97,100)]

## CD.Block\_Code Level Name VCat SexRatio  
## 1 244 VILLAGE Fursungi V4 901.6667

## End of File