Data_Analytics_Assgn_AndhraPradesh_gov

Tushar H J

February 15, 2018

Loading Required Libraries

```
library(dplyr)

library(ggplot2)
library(openxlsx)
library(stringr)
```

Reading Data from Excel sheet (assignment.xlxs)

```
Master_sheet <- read.xlsx("C:/Users/user/Desktop/Data Science/Projects/Internshala/Andhra Pradesh/assignment.xlsx", sheet = 1)
Data_sheet <- read.xlsx("C:/Users/user/Desktop/Data Science/Projects/Internshala/Andhra Pradesh/assignment.xlsx", sheet = 2)
```

checking summary of data

```
(head(Master_sheet))

(summary(Master_sheet))

(colnames(Master_sheet))

##  [1] "DISTRICT.NAME"     "MANDAL.NAME"       "VILLAGE.NAME"
##  [4] "SCHOOL.UDISE.CODE" "SCHOOL.NAME"       "ADDRESS"
##  [7] "AREA"              "SCHOOL.CATEGORY"   "SCHOOL.MANGEMENT"
## [10] "Enrolment"

(head(Data_sheet))
```

Checking missing values in data

```
sapply(Data_sheet,function(x)sum(is.na(x)))

##          District            Mandal School.UDISE.Code       School.Name
##                 0                 0                 0                 1
##        Enrollment
##               204
```

```r
sapply(Master_sheet,function(x)sum(is.na(x)))
```

```
##     DISTRICT.NAME       MANDAL.NAME      VILLAGE.NAME SCHOOL.UDISE.CODE
##                 0                 0                 1                 0
##       SCHOOL.NAME           ADDRESS              AREA   SCHOOL.CATEGORY
##                 4              1682                 0                 0
##  SCHOOL.MANGEMENT         Enrolment
##                 0                 0
```

converting all district names to upper case

```r
Data_sheet$District <- toupper(Data_sheet$District)
table(Data_sheet$District)
```

```
##
##     ANANTAPUR    ANANTHAPUR      CHITTOOR     CHITTOOR      CHITTOROR
##             4             2             2             1             1
## EAST GODAVARI        GUNTUR        KADAPA       KRISHNA       KURNOOL
##             7            37             2            64             6
##       NELLORE      PRAKASAM    SRIKAKULAM VISAKHAPATNAM  VIZIANAGARAM
##             3            26             2            28             4
## WEST GODAVARI
##            15
```

ANANTAPUR and ANANTHAPUR are same –> ANANTAPUR

CHITTOOR and CHITTOROR are same –> CHITTOOR

```r
Data_sheet$District <- (str_trim(Data_sheet$District)) # removing whitespaces

r = Data_sheet[, "District"] == "ANANTHAPUR"
Data_sheet[r,"District"] = "ANANTAPUR"

r = Data_sheet[, "District"] == "CHITTOROR"
Data_sheet[r,"District"] = "CHITTOOR"

table(Data_sheet$District)
```

```
##
##     ANANTAPUR      CHITTOOR EAST GODAVARI        GUNTUR        KADAPA
##             6             4             7            37             2
##       KRISHNA       KURNOOL       NELLORE      PRAKASAM    SRIKAKULAM
##            64             6             3            26             2
## VISAKHAPATNAM  VIZIANAGARAM WEST GODAVARI
##            28             4            15
```

1. Compare Master Sheet with Data Sheet in the Assignment Work Book take UDISE Code

as unique code for every school and fill enrollment for Data Sheet from Master Sheet

```
COUNT = 1
for (x in Data_sheet$School.UDISE.Code){
r =  Master_sheet[,"SCHOOL.UDISE.CODE"] == x
Data_sheet[COUNT,"Enrollment"] <- Master_sheet[r,"Enrolment"]
COUNT = COUNT +1
}
(head(Data_sheet$Enrollment))
```

```
## [1] 157 241 111  97 239 488
```

2.Compare Master Sheet with Data Sheet in the Assignment Work Book take UDISE Code as

unique code for every school and fill School Category for Data Sheet from Master Sheet

```
COUNT = 1
for(x in Data_sheet$School.UDISE.Code){
  r = Master_sheet[,"SCHOOL.UDISE.CODE"] == x
  Data_sheet[COUNT, "School_Category"] <- Master_sheet[r, "SCHOOL.CATEGORY"]
  COUNT = COUNT +1
}
colnames(Data_sheet)
```

```
## [1] "District"          "Mandal"            "School.UDISE.Code"
## [4] "School.Name"       "Enrollment"        "School_Category"
```

3. Visualize the DATA Sheet portraying Parameters like how many schools per District, Total

Impact in terms of enrollment per district, How many Primary , how many primary with Upper Primary,

how many upper primary with secondary schools overall (School Category)

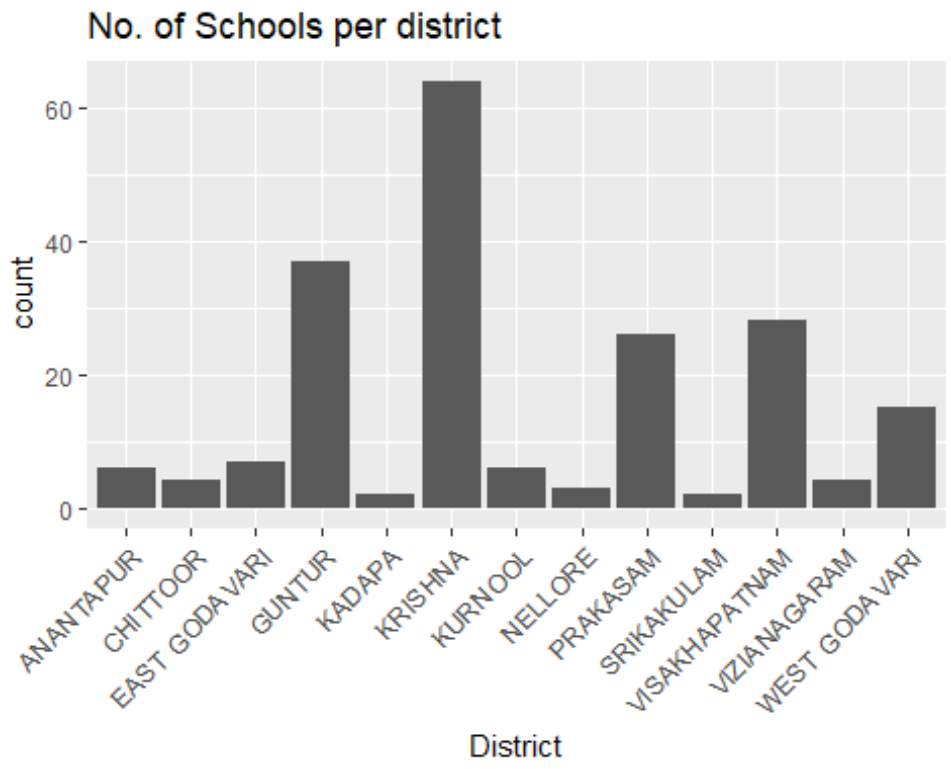
Number of schools per district

```
Schools_District <- as.factor(Data_sheet$District)
School_count_district <- Data_sheet %>% group_by(District) %>% summarise(Coun
t = n())
print(School_count_district)
```

```
## # A tibble: 13 x 2
##    District      Count
##    <chr>         <int>
##  1 ANANTAPUR         6
##  2 CHITTOOR          4
##  3 EAST GODAVARI     7
##  4 GUNTUR           37
##  5 KADAPA            2
##  6 KRISHNA          64
##  7 KURNOOL           6
##  8 NELLORE           3
##  9 PRAKASAM         26
## 10 SRIKAKULAM        2
## 11 VISAKHAPATNAM    28
## 12 VIZIANAGARAM      4
## 13 WEST GODAVARI    15
```

Plot for No. of Schools per district

```
ggplot(data = Data_sheet, aes(x = Data_sheet$District)) +
  xlab("District")+
  ggtitle("No. of Schools per district") +
  stat_count() +
  theme(axis.text.x = element_text(size  = 10,angle = 45, hjust = 1,vjust = 1
))
```



Total Impact in terms of enrollment per district

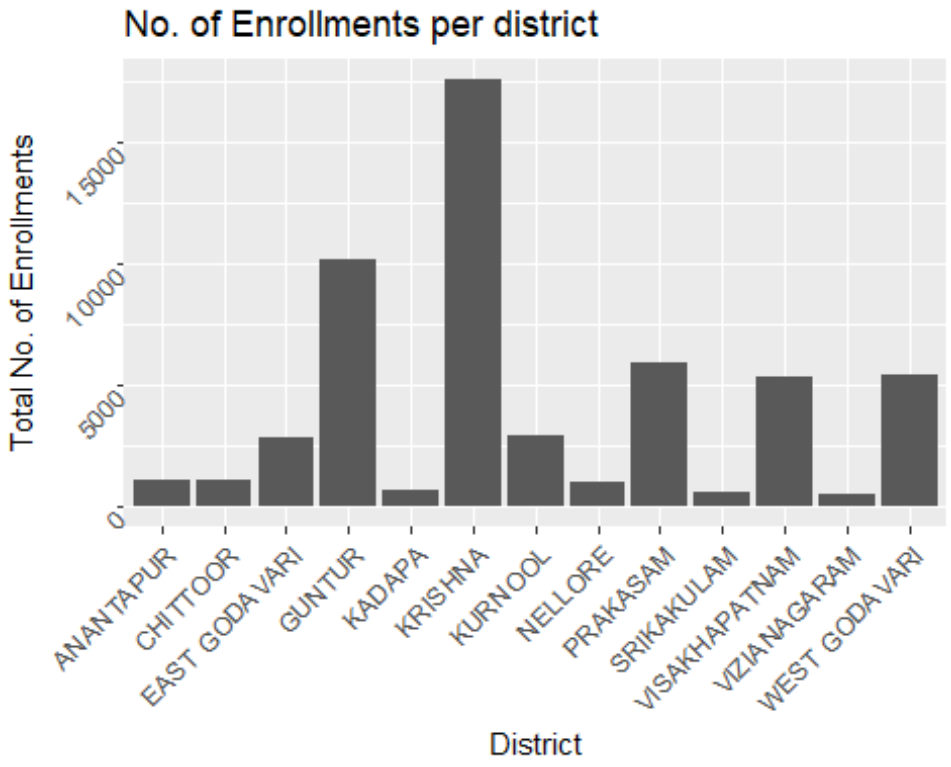
```
Enrol_District <- Data_sheet %>%
                  group_by(District) %>%
                  summarise(Total_Enrollments = sum(Enrollment))

Enrol_District = as.data.frame(Enrol_District)
(Enrol_District)
```

```
##          District Total_Enrollments
## 1       ANANTAPUR              1077
## 2        CHITTOOR              1046
## 3  EAST GODAVARI              2796
## 4          GUNTUR             10112
## 5          KADAPA               669
## 6         KRISHNA             17555
## 7         KURNOOL              2910
## 8         NELLORE               969
## 9        PRAKASAM              5898
## 10    SRIKAKULAM               537
## 11 VISAKHAPATNAM              5335
## 12  VIZIANAGARAM               463
## 13 WEST GODAVARI              5356
```

Plot of Enrollments for each district

```
ggplot(data = Enrol_District, aes(x=factor(Enrol_District$District), y = Enro
l_District$Total_Enrollments))+
  geom_bar(stat = 'identity')+
  xlab("District")+
  ylab("Total No. of Enrollments")+
  ggtitle("No. of Enrollments per district") +
  theme(axis.text = element_text(size = 10, angle = 45, hjust = 1, vjust = 1)
)
```



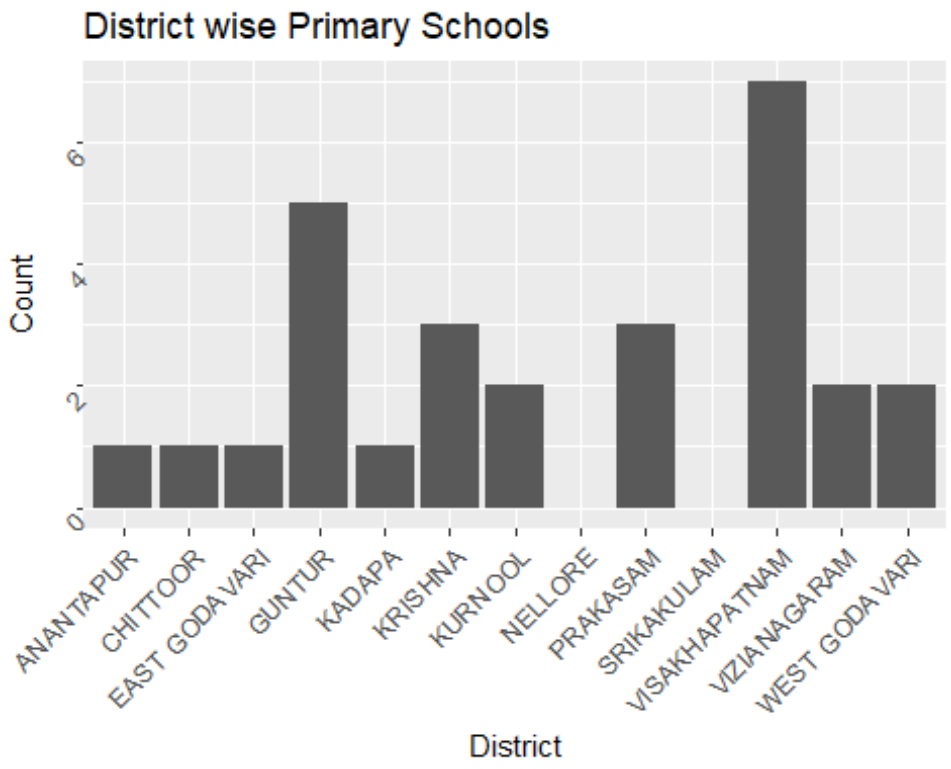
Districtwise School categories

```
(School_cat_district <- Data_sheet %>%
                        group_by(District) %>%
                        summarise(Primary = sum(School_Category == '1-Primary'
),
                         Primary_Upper_Primary = sum(School_Category == "2-Pri
mary with Upper Primary"),
                         Upper_Prim_Secondary = sum(School_Category=="7-Upper
Pr. and Secondary")                          )
  )
```

```
## # A tibble: 13 x 4
##    District      Primary Primary_Upper_Primary Upper_Prim_Secondary
##    <chr>           <int>                 <int>                <int>
##  1 ANANTAPUR           1                     0                    5
##  2 CHITTOOR            1                     0                    3
##  3 EAST GODAVARI       1                     1                    5
##  4 GUNTUR              5                     0                   32
##  5 KADAPA              1                     0                    1
##  6 KRISHNA             3                     2                   59
##  7 KURNOOL             2                     0                    4
##  8 NELLORE             0                     0                    3
##  9 PRAKASAM            3                     0                   23
## 10 SRIKAKULAM          0                     0                    2
## 11 VISAKHAPATNAM       7                     7                   14
## 12 VIZIANAGARAM        2                     0                    2
## 13 WEST GODAVARI       2                     0                   13
```

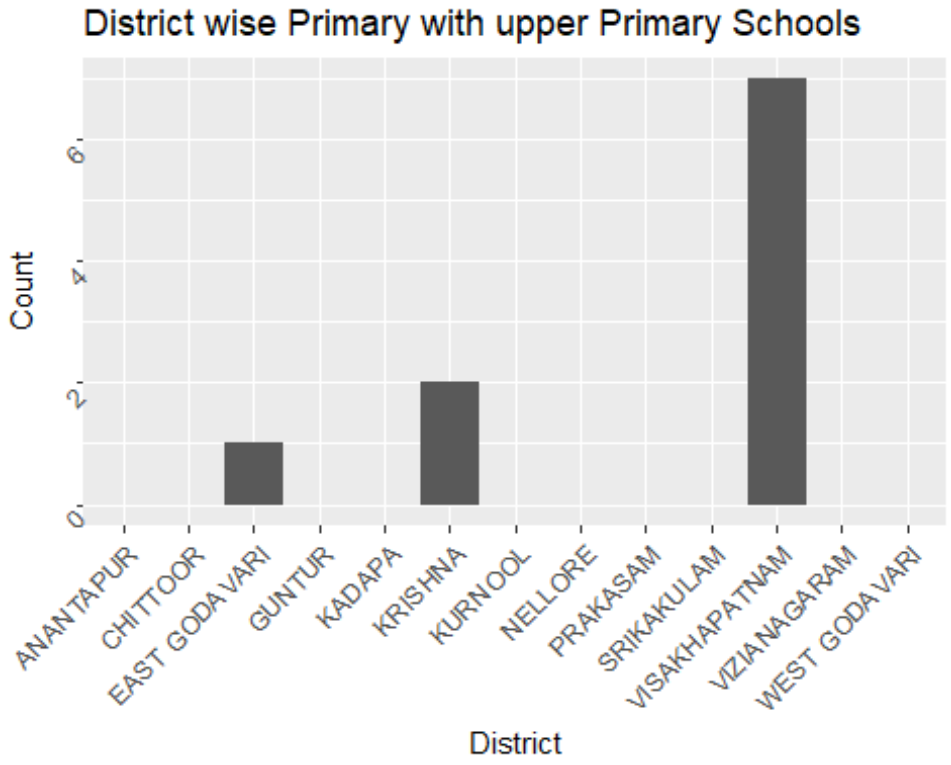
Plot for District wise Primary Schools

```
ggplot(data = School_cat_district, aes(x=factor(School_cat_district$District)
, y = School_cat_district$Primary))+
  geom_bar(stat = 'identity')+
  xlab("District")+
  ylab("Count")+
  ggtitle("District wise Primary Schools") +
  theme(axis.text = element_text(size = 10, angle = 45, hjust = 1, vjust = 1)
)
```



Plot for District wise Primary with upper Primary Schools

```
ggplot(data = School_cat_district, aes(x=factor(School_cat_district$District)
, y = School_cat_district$Primary_Upper_Primary))+
  geom_bar(stat = 'identity')+
  xlab("District")+
  ylab("Count")+
  ggtitle("District wise Primary with upper Primary Schools") +
  theme(axis.text = element_text(size = 10, angle = 45, hjust = 1, vjust =1))
```



#Plot for District wise Upper Primary with Secondary Schools

```r
ggplot(data = School_cat_district, aes(x=factor(School_cat_district$District)
, y = School_cat_district$Upper_Prim_Secondary))+
  geom_bar(stat = 'identity')+
  xlab("District")+
  ylab("Count")+
  ggtitle("District wise Upper Primary with Secondary Schools") +
  theme(axis.text = element_text(size = 10, angle = 45, hjust = 1, vjust =1))
```