Surname Analysis of Munger LokSabha electoral data 2019

Our focus will be on 5 constituency of Munger i.e,

1. Munger

2. Jamalpur

3. Lakhisarai

4. Mokama

5. Badh

We are provided the data in the .bak format. So, we extracted the excel file from the SQL server.

Since excel files takes a lot of time to load in R stdio so, we converted it to the csv file using online excel to csv converter.

Loading the required libraries

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.6
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(humaniformat)
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:purrr':
##
##     set_names
```

```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
library(tools)
```

Reading the data

```
df_munger=read.csv("D:/Janta ka mood Intern/Munger_Constituency_2019_electoral/Munger_165.csv")
df_jamalpur=read.csv("D:/Janta ka mood Intern/Munger_Constituency_2019_electoral/Jamalpur_166.csv")
df_lakhisarai=read.csv("D:/Janta ka mood Intern/Munger_Constituency_2019_electoral/Lakhisarai_168.csv")
df_mokama=read.csv("D:/Janta ka mood Intern/Munger_Constituency_2019_electoral/Mokama_178.csv")
df_badh=read.csv("D:/Janta ka mood Intern/Munger_Constituency_2019_electoral/Badh_179.csv")
```

Selecting the feature columns of Munger

```
colnames(df_munger)
```

```
##  [1] "X"                     "ACNo"
##  [3] "PartNo"                "SectionNo"
##  [5] "SNo"                   "HouseNoEn"
##  [7] "HouseNo"               "VoterNameEn"
##  [9] "VoterName"             "Sex"
## [11] "RelationNameEn"        "RelationName"
## [13] "RelationType"          "Age"
## [15] "VoterID"               "StatusType"
## [17] "ContactNo"             "SectionName"
## [19] "SectionNameEn"         "ACName"
## [21] "ACNameEn"              "PollingStationAddressEn"
## [23] "PollingStationAddress" "DistrictNameEn"
## [25] "DistrictName"          "X.1"
## [27] "X.2"
```

```
df_munger <- df_munger %>% select(PartNo,SectionNo,HouseNo,VoterNameEn,Sex,Age)
df_munger <- df_munger %>% filter(Sex=="M",!is.na(VoterNameEn))
head(df_munger)
```

```
##   PartNo SectionNo HouseNo         VoterNameEn Sex Age
## 1    199         1      87 Afisar Prasad Yadav   M  48
## 2    199         1      87 Dhiraj Prasad Yadav   M  42
## 3    199         1      88        Suresh Yadav   M  72
## 4    199         1      89 Mahesh Prasad Yadav   M  67
## 5    199         1      89          Sanjiv Ray   M  37
## 6    199         1      89        Ravindra Ray   M  54
```

selecting the feature columns of Jamalpur

```
colnames(df_jamalpur)
```

```
##  [1] "X"                     "ACNo"
##  [3] "PartNo"                "SectionNo"
##  [5] "SNo"                   "HouseNoEn"
##  [7] "HouseNo"               "VoterNameEn"
##  [9] "VoterName"             "Sex"
## [11] "RelationNameEn"        "RelationName"
## [13] "RelationType"          "Age"
## [15] "VoterID"               "StatusType"
## [17] "ContactNo"             "SectionName"
## [19] "SectionNameEn"         "ACName"
## [21] "ACNameEn"              "PollingStationAddressEn"
## [23] "PollingStationAddress" "DistrictNameEn"
## [25] "DistrictName"          "X.1"
## [27] "X.2"
```

```
df_jamalpur <- df_jamalpur %>% select(PartNo, SectionNo, HouseNo, VoterNameEn, Sex, Age)
df_jamalpur <- df_jamalpur %>% filter(Sex=="M",!is.na(VoterNameEn))
head(df_jamalpur)
```

```
##   PartNo SectionNo HouseNo           VoterNameEn Sex Age
## 1      1         1      14          Anandi Malik   M  40
## 2      1         1      14             anil sada   M  30
## 3      1         1      14        ravindra malik   M  29
## 4      1         1      15 Rajendra Prasad Mahato   M  72
## 5      1         1      15    Inndubhushan Mehata   M  63
## 6      1         1      15          Sandip Patil   M  32
```

selecting the feature columns of Lakhisarai

```
colnames(df_lakhisarai)
```

```
##  [1] "X"                      "ACNo"
##  [3] "PartNo"                 "SectionNo"
##  [5] "SNo"                    "HouseNoEn"
##  [7] "HouseNo"                "VoterNameEn"
##  [9] "VoterName"              "Sex"
## [11] "RelationNameEn"         "RelationName"
## [13] "RelationType"           "Age"
## [15] "VoterID"                "StatusType"
## [17] "ContactNo"              "SectionName"
## [19] "SectionNameEn"          "ACName"
## [21] "ACNameEn"               "PollingStationAddressEn"
## [23] "PollingStationAddress"  "DistrictNameEn"
## [25] "DistrictName"           "X.1"
## [27] "X.2"                    "X.3"
## [29] "X.4"
```

```
df_lakhisarai <- df_lakhisarai %>% select(PartNo, SectionNo, HouseNo, VoterNameEn, Sex, Age)
df_lakhisarai <- df_lakhisarai %>% filter(Sex=="M",!is.na(VoterNameEn))
head(df_lakhisarai)
```

```
##   PartNo SectionNo HouseNo            VoterNameEn Sex Age
## 1    139         1     102     Ganesh Kumar Varma   M  48
## 2    139         1     102        UMESH KUMAR SAW   M  42
## 3    139         1     102       AMIT KUMAR VERMA   M  27
## 4    139         1     102             Ajit Verma   M  28
## 5    139         1     102 Mahendra Parshad Verma   M  68
## 6    139         1     103       Suman Prao Verma   M  30
```

selecting the feature columns of Mokama

```
colnames(df_mokama)
```

```
##  [1] "ACNo"                   "PartNo"
##  [3] "SectionNo"              "SNo"
##  [5] "HouseNoEn"              "HouseNo"
##  [7] "VoterNameEn"            "VoterName"
##  [9] "Sex"                    "RelationNameEn"
## [11] "RelationName"           "RelationType"
## [13] "Age"                    "VoterID"
## [15] "StatusType"             "ContactNo"
## [17] "SectionName"            "SectionNameEn"
## [19] "ACName"                 "ACNameEn"
## [21] "PollingStationAddressEn" "PollingStationAddress"
## [23] "DistrictNameEn"         "DistrictName"
## [25] "Column1"
```

```
df_mokama <- df_mokama %>% select(PartNo, SectionNo, HouseNo, VoterNameEn, Sex, Age)
df_mokama <- df_mokama %>% filter(Sex=="M",!is.na(VoterNameEn))
head(df_mokama)
```

```
##   PartNo SectionNo HouseNo       VoterNameEn Sex Age
## 1     52         1       2   Mahendra Prasad   M  54
## 2     52         1       2      Vikas Kumar    M  31
## 3     52         1       2     Bhavesh Kumar   M  37
## 4     52         1       2      Vikash Kumar   M  24
## 5     52         1       2    Abhishek Kumar   M  24
## 6     52         1       2 CHANDRAMANI KUMAR   M  21
```

selecting the feature columns of Badh

```
colnames(df_badh)
```

```
##  [1] "X"                     "ACNo"
##  [3] "PartNo"                "SectionNo"
##  [5] "SNo"                   "HouseNoEn"
##  [7] "HouseNo"               "VoterNameEn"
##  [9] "VoterName"             "Sex"
## [11] "RelationNameEn"        "RelationName"
## [13] "RelationType"          "Age"
## [15] "VoterID"               "StatusType"
## [17] "ContactNo"             "SectionName"
## [19] "SectionNameEn"         "ACName"
## [21] "ACNameEn"              "PollingStationAddressEn"
## [23] "PollingStationAddress" "DistrictNameEn"
## [25] "DistrictName"          "X.1"
## [27] "X.2"
```

```
df_badh <- df_badh %>% select(PartNo, SectionNo, HouseNo, VoterNameEn, Sex, Age)
df_badh <- df_badh %>% filter(Sex=="M",!is.na(VoterNameEn))
head(df_badh)
```

```
##   PartNo SectionNo HouseNo   VoterNameEn Sex Age
## 1      1         2      27 Jitesh Kumar    M  38
## 2      1         2      27 Rakesh Kumar    M  40
## 3      1         2      27   Rahul Kumar   M  27
## 4      1         2      27  SUDHU PASWAN   M  25
## 5      1         2      27    ARUN MAHTO   M  26
## 6      1         2      27 GORAKH MAHATO   M  29
```

Extracting the last names of Munger

```
df_munger <- df_munger %>% add_column("FirstName"=NA,"MiddleName"=NA,"LastName"=NA)
nam_munger <- parse_names(df_munger$VoterNameEn)
df_munger$FirstName <- nam_munger$first_name
df_munger$MiddleName <- nam_munger$middle_name
df_munger$LastName <- nam_munger$last_name
df_munger$LastName <- toTitleCase(tolower(df_munger$LastName))
head(df_munger)
```

```
##   PartNo SectionNo HouseNo         VoterNameEn Sex Age FirstName MiddleName
## 1    199         1      87 Afisar Prasad Yadav   M  48    Afisar     Prasad
## 2    199         1      87 Dhiraj Prasad Yadav   M  42    Dhiraj     Prasad
## 3    199         1      88        Suresh Yadav   M  72    Suresh       <NA>
## 4    199         1      89 Mahesh Prasad Yadav   M  67    Mahesh     Prasad
## 5    199         1      89          Sanjiv Ray   M  37    Sanjiv       <NA>
## 6    199         1      89        Ravindra Ray   M  54  Ravindra       <NA>
##   LastName
## 1    Yadav
## 2    Yadav
## 3    Yadav
## 4    Yadav
## 5      Ray
## 6      Ray
```

Extracting the last names of Jamalpur

```
df_jamalpur <- df_jamalpur %>% add_column("FirstName"=NA,"MiddleName"=NA,"LastName"=NA)
nam_jamalpur <- parse_names(df_jamalpur$VoterNameEn)
df_jamalpur$FirstName <- nam_jamalpur$first_name
df_jamalpur$MiddleName <- nam_jamalpur$middle_name
df_jamalpur$LastName <- nam_jamalpur$last_name
df_jamalpur$LastName <- toTitleCase(tolower(df_jamalpur$LastName))
head(df_jamalpur)
```

```
##   PartNo SectionNo HouseNo            VoterNameEn Sex Age    FirstName
## 1      1         1      14           Anandi Malik   M  40       Anandi
## 2      1         1      14              anil sada   M  30         anil
## 3      1         1      14         ravindra malik   M  29     ravindra
## 4      1         1      15 Rajendra Prasad Mahato   M  72     Rajendra
## 5      1         1      15    Inndubhushan Mehata   M  63 Inndubhushan
## 6      1         1      15           Sandip Patil   M  32       Sandip
##   MiddleName LastName
## 1       <NA>    Malik
## 2       <NA>     Sada
## 3       <NA>    Malik
## 4     Prasad   Mahato
## 5       <NA>   Mehata
## 6       <NA>    Patil
```

Extracting the last names of Lakhisarai

```
df_lakhisarai <- df_lakhisarai %>% add_column("FirstName"=NA,"MiddleName"=NA,"LastName"=NA)
nam_lakhisarai <- parse_names(df_lakhisarai$VoterNameEn)
df_lakhisarai$FirstName <- nam_lakhisarai$first_name
df_lakhisarai$MiddleName <- nam_lakhisarai$middle_name
df_lakhisarai$LastName <- nam_lakhisarai$last_name
df_lakhisarai$LastName <- toTitleCase(tolower(df_lakhisarai$LastName))
head(df_lakhisarai)
```

```
##   PartNo SectionNo HouseNo            VoterNameEn Sex Age FirstName MiddleName
## 1    139         1     102     Ganesh Kumar Varma   M  48    Ganesh      Kumar
## 2    139         1     102        UMESH KUMAR SAW   M  42     UMESH      KUMAR
## 3    139         1     102       AMIT KUMAR VERMA   M  27      AMIT      KUMAR
## 4    139         1     102             Ajit Verma   M  28      Ajit       <NA>
## 5    139         1     102 Mahendra Parshad Verma   M  68  Mahendra    Parshad
## 6    139         1     103       Suman Prao Verma   M  30     Suman       Prao
##   LastName
## 1    Varma
## 2      Saw
## 3    Verma
## 4    Verma
## 5    Verma
## 6    Verma
```

Extracting the last names of Mokama

```
df_mokama <- df_mokama %>% add_column("FirstName"=NA,"MiddleName"=NA,"LastName"=NA)
nam_mokama <- parse_names(df_mokama$VoterNameEn)
df_mokama$FirstName <- nam_mokama$first_name
df_mokama$MiddleName <- nam_mokama$middle_name
df_mokama$LastName <- nam_mokama$last_name
df_mokama$LastName <- toTitleCase(tolower(df_mokama$LastName))
head(df_mokama)
```

```
##   PartNo SectionNo HouseNo       VoterNameEn Sex Age   FirstName MiddleName
## 1     52         1       2   Mahendra Prasad   M  54    Mahendra       <NA>
## 2     52         1       2      Vikas Kumar    M  31       Vikas      Kumar
## 3     52         1       2     Bhavesh Kumar   M  37     Bhavesh       <NA>
## 4     52         1       2      Vikash Kumar   M  24      Vikash       <NA>
## 5     52         1       2    Abhishek Kumar   M  24    Abhishek       <NA>
## 6     52         1       2 CHANDRAMANI KUMAR   M  21 CHANDRAMANI       <NA>
##   LastName
## 1   Prasad
## 2     <NA>
## 3    Kumar
## 4    Kumar
## 5    Kumar
## 6    Kumar
```

Extracting the last names of Badh

```
df_badh <- df_badh %>% add_column("FirstName"=NA,"MiddleName"=NA,"LastName"=NA)
nam_badh <- parse_names(df_badh$VoterNameEn)
df_badh$FirstName <- nam_badh$first_name
df_badh$MiddleName <- nam_badh$middle_name
df_badh$LastName <- nam_badh$last_name
df_badh$LastName <- toTitleCase(tolower(df_badh$LastName))
head(df_badh)
```

```
##   PartNo SectionNo HouseNo   VoterNameEn Sex Age FirstName MiddleName LastName
## 1      1         2      27 Jitesh Kumar    M  38    Jitesh      Kumar     <NA>
## 2      1         2      27 Rakesh Kumar    M  40    Rakesh      Kumar     <NA>
## 3      1         2      27   Rahul Kumar   M  27     Rahul       <NA>    Kumar
## 4      1         2      27  SUDHU PASWAN   M  25     SUDHU       <NA>   Paswan
## 5      1         2      27    ARUN MAHTO   M  26      ARUN       <NA>    Mahto
## 6      1         2      27 GORAKH MAHATO   M  29    GORAKH       <NA>   Mahato
```

Selecting the head of the family and counting the surname of munger

```
df_munger <- df_munger %>% filter(! LastName  %in% c("Kumar",NA))
df2_munger <- df_munger %>% group_by(PartNo,SectionNo,HouseNo) %>% top_n(1,Age)
head(df2_munger)
```

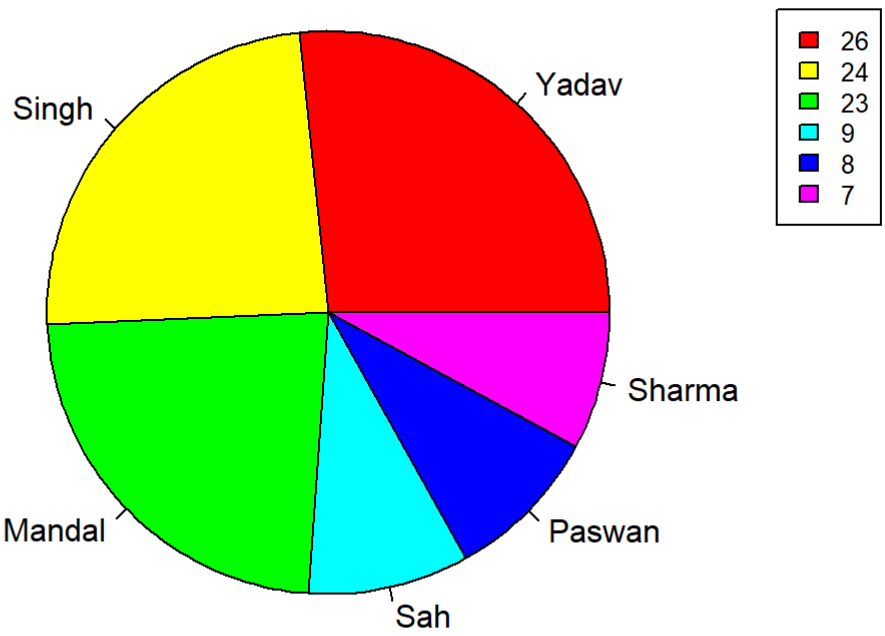
```
## # A tibble: 6 x 9
## # Groups:   PartNo, SectionNo, HouseNo [6]
##   PartNo SectionNo HouseNo VoterNameEn Sex   Age   FirstName MiddleName LastName
##    <int>     <int> <chr>   <chr>       <chr> <chr> <chr>     <chr>      <chr>
## 1    199         1 87      Afisar Pra~ M     48    Afisar    Prasad     Yadav
## 2    199         1 88      Suresh Yad~ M     72    Suresh    <NA>       Yadav
## 3    199         1 89      Mahesh Pra~ M     67    Mahesh    Prasad     Yadav
## 4    199         1 90      Ashok Ray   M     49    Ashok     <NA>       Ray
## 5    199         1 91      Samadan Ya~ M     89    Samadan   <NA>       Yadav
## 6    199         1 92      Pradip Yad~ M     61    Pradip    <NA>       Yadav
```

```
df2_munger <- df2_munger %>% group_by(LastName) %>% mutate(count=n())
surname_munger <- df2_munger %>% select(LastName,count)
surname_munger <- unique(surname_munger)
surname_munger<- surname_munger %>% arrange(desc(count))
head(surname_munger)
```

```
## # A tibble: 6 x 2
## # Groups:   LastName [6]
##   LastName count
##   <chr>    <int>
## 1 Yadav     5406
## 2 Singh     4872
## 3 Mandal    4706
## 4 Sah       1880
## 5 Paswan    1814
## 6 Sharma    1608
```

```
top_surname_munger <- head(surname_munger)

label <- top_surname_munger$LastName[1:6]
per <- top_surname_munger$count[1:6]/sum(top_surname_munger$count[1:6])*100
per <- as.integer(per)
pie(top_surname_munger$count[1:6],labels =label ,radius=1,main="Top 6 surnames of MUNGER",col
=rainbow(6))
legend("topright",inset=0.001,legend=per,cex=0.8, fill = rainbow(6))
```



Selecting the head of the family and counting the surname of jamalpur

```
df_jamalpur <- df_jamalpur %>% filter(! LastName  %in% c("Kumar",NA))
df2_jamalpur<- df_jamalpur %>% group_by(PartNo,SectionNo,HouseNo) %>% top_n(1,Age)
head(df2_jamalpur)
```

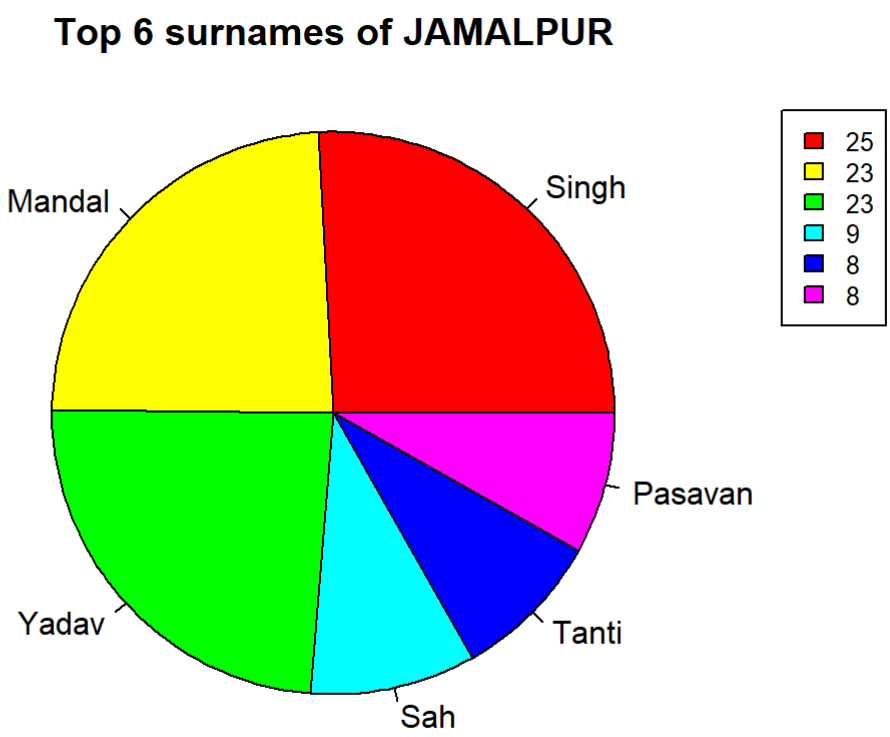
```
## # A tibble: 6 x 9
## # Groups:   PartNo, SectionNo, HouseNo [6]
##   PartNo SectionNo HouseNo VoterNameEn Sex   Age   FirstName MiddleName LastName
##   <chr>  <chr>     <chr>   <chr>       <chr> <chr> <chr>     <chr>      <chr>
## 1 1      1         15      Rajendra P~ M     72    Rajendra  Prasad     Mahato
## 2 1      1         16      Subodh Pas~ M     52    Subodh    <NA>       Pasavan
## 3 1      1         17      Umesh Sav   M     50    Umesh     <NA>       Sav
## 4 1      1         18      Dinesh Tha~ M     38    Dinesh    <NA>       Thakur
## 5 1      1         19      Inasu Miya  M     65    Inasu     <NA>       Miya
## 6 1      1         20      Ala Uddin   M     48    Ala       <NA>       Uddin
```

```
df2_jamalpur <- df2_jamalpur %>% group_by(LastName) %>% mutate(count=n())
surname_jamalpur <- df2_jamalpur %>% select(LastName,count)
surname_jamalpur <- unique(surname_jamalpur)
surname_jamalpur<- surname_jamalpur %>% arrange(desc(count))
head(surname_jamalpur)
```

```
## # A tibble: 6 x 2
## # Groups:   LastName [6]
##   LastName count
##   <chr>    <int>
## 1 Singh     5392
## 2 Mandal    5000
## 3 Yadav     4982
## 4 Sah       1980
## 5 Tanti     1795
## 6 Pasavan   1704
```

```
top_surname_jamalpur <- head(surname_jamalpur)

label <- top_surname_jamalpur$LastName[1:6]
per <- top_surname_jamalpur$count[1:6]/sum(top_surname_jamalpur$count[1:6])*100
per <- as.integer(per)
pie(top_surname_jamalpur$count[1:6],labels =label ,radius=1,main="Top 6 surnames of JAMALPUR"
,col=rainbow(6))
legend("topright",inset=0.001,legend=per,cex=0.8, fill = rainbow(6))
```



Selecting the head of the family and counting the surname of lakhisarai

```
df_lakhisarai <- df_lakhisarai %>% filter(! LastName  %in% c("Kumar",NA))
df2_lakhisarai <- df_lakhisarai %>% group_by(PartNo,SectionNo,HouseNo) %>% top_n(1,Age)
head(df2_lakhisarai)
```

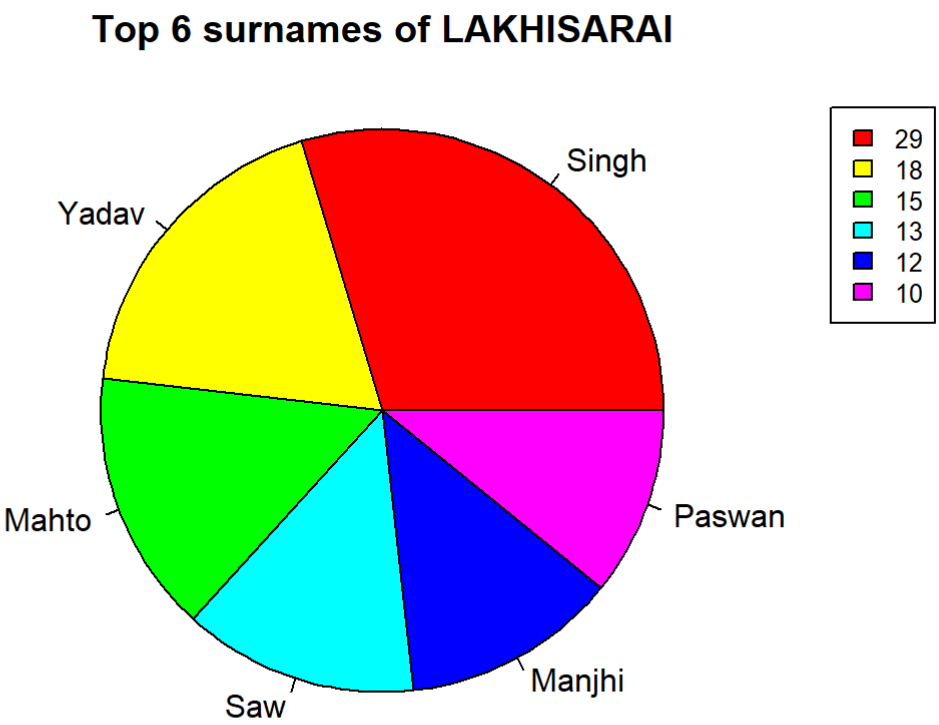
```
## # A tibble: 6 x 9
## # Groups:   PartNo, SectionNo, HouseNo [6]
##   PartNo SectionNo HouseNo VoterNameEn Sex   Age   FirstName MiddleName LastName
##   <chr>  <chr>     <chr>   <chr>       <chr> <chr> <chr>     <chr>      <chr>
## 1 139    1         103     Suman Prao~ M     30    Suman     Prao       Verma
## 2 139    1         104     Manoj Ku S~ M     44    Manoj     Ku         Saw
## 3 139    1         105     Chaitu SAW  M     99    Chaitu    <NA>       Saw
## 4 139    1         106     Aditya Kum~ M     47    Aditya    Kumar      Saw
## 5 139    1         107     MOHAMMAD I~ M     76    MOHAMMAD  <NA>       Idaris
## 6 139    1         110     Kisun Das   M     74    Kisun     <NA>       Das
```

```
df2_lakhisarai <- df2_lakhisarai %>% group_by(LastName) %>% mutate(count=n())
surname_lakhisarai <- df2_lakhisarai %>% select(LastName,count)
surname_lakhisarai <- unique(surname_lakhisarai)
surname_lakhisarai<- surname_lakhisarai %>% arrange(desc(count))
head(surname_lakhisarai)
```

```
## # A tibble: 6 x 2
## # Groups:   LastName [6]
##   LastName count
##   <chr>    <int>
## 1 Singh     7052
## 2 Yadav     4409
## 3 Mahto     3587
## 4 Saw       3223
## 5 Manjhi    2949
## 6 Paswan    2580
```

```
top_surname_lakhisarai <- head(surname_lakhisarai)

label <- top_surname_lakhisarai$LastName[1:6]
per <- top_surname_lakhisarai$count[1:6]/sum(top_surname_lakhisarai$count[1:6])*100
per <- as.integer(per)
pie(top_surname_lakhisarai$count[1:6],labels =label ,radius=1,main="Top 6 surnames of LAKHISA
RAI",col=rainbow(6))
legend("topright",inset=0.001,legend=per,cex=0.8, fill = rainbow(6))
```



Selecting the head of the family and counting the surname of mokama

```
df_mokama <- df_mokama %>% filter(! LastName  %in% c("Kumar",NA))
df2_mokama <- df_mokama %>% group_by(PartNo,SectionNo,HouseNo) %>% top_n(1,Age)
head(df2_mokama)
```

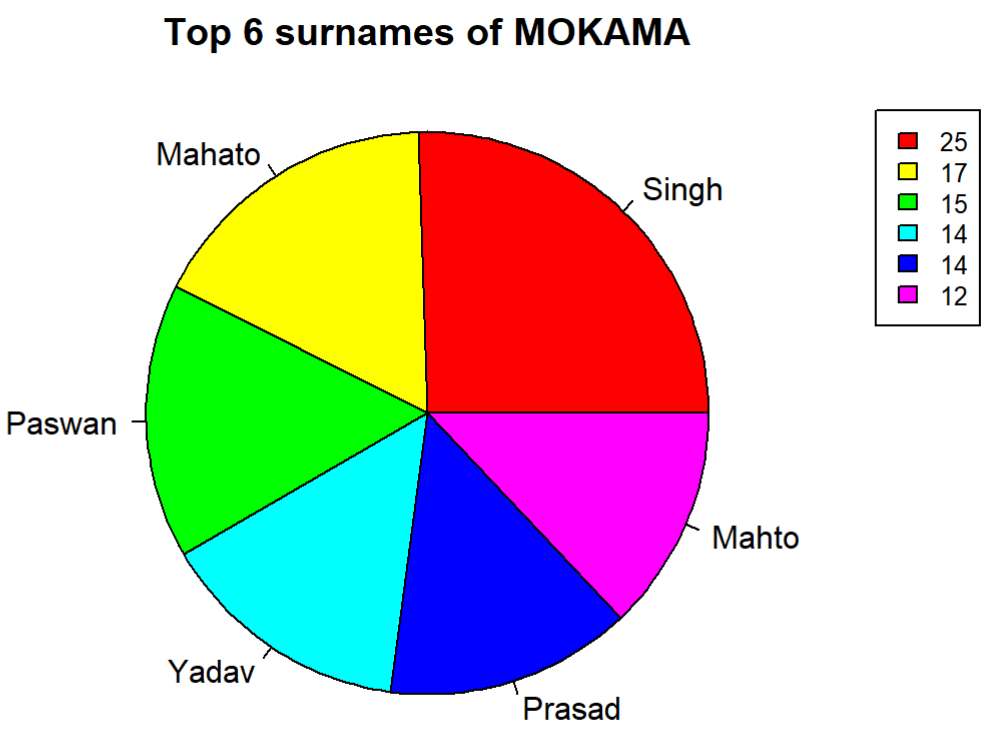
```
## # A tibble: 6 x 9
## # Groups:   PartNo, SectionNo, HouseNo [6]
##   PartNo SectionNo HouseNo VoterNameEn Sex   Age   FirstName MiddleName LastName
##   <chr>  <chr>     <chr>   <chr>       <chr> <chr> <chr>     <chr>      <chr>
## 1 52     1         3       Vishwanath~ M     66    Vishwana~ Pra        Singh
## 2 2      1         55      Mahendra D~ M     72    Mahendra  <NA>       Das
## 3 2      1         56      RAMESHWAR ~ M     85    RAMESHWAR <NA>       Das
## 4 2      1         57      Ramsharan ~ M     O     Ramsharan <NA>       Yadav
## 5 2      1         58      Krishna Das M     73    Krishna   <NA>       Das
## 6 2      1         59      Ramanandan~ M     74    Ramanand~ Prasad     Yadav
```

```
df2_mokama <- df2_mokama %>% group_by(LastName) %>% mutate(count=n())
surname_mokama <- df2_mokama %>% select(LastName,count)
surname_mokama <- unique(surname_mokama)
surname_mokama<- surname_mokama %>% arrange(desc(count))
head(surname_mokama)
```

```
## # A tibble: 6 x 2
## # Groups:   LastName [6]
##   LastName count
##   <chr>    <int>
## 1 Singh     3336
## 2 Mahato    2232
## 3 Paswan    2061
## 4 Yadav     1902
## 5 Prasad    1852
## 6 Mahto     1694
```

```
top_surname_mokama <- head(surname_mokama)

label <- top_surname_mokama$LastName[1:6]
per <- top_surname_mokama$count[1:6]/sum(top_surname_mokama$count[1:6])*100
per <- as.integer(per)
pie(top_surname_mokama$count[1:6],labels =label ,radius=1,main="Top 6 surnames of MOKAMA",col
=rainbow(6))
legend("topright",inset=0.001,legend=per,cex=0.8, fill = rainbow(6))
```



Selecting the head of the family and counting the surname of badh

```
df_badh <- df_badh %>% filter(! LastName  %in% c("Kumar",NA))
df2_badh <- df_badh %>% group_by(PartNo,SectionNo,HouseNo) %>% top_n(1,Age)
head(df2_badh)
```

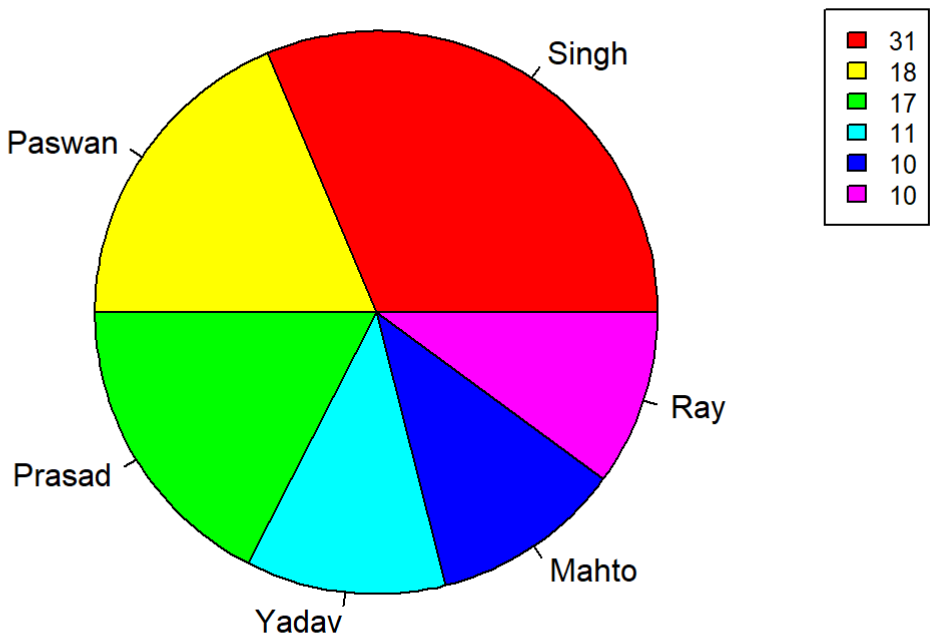
```
## # A tibble: 6 x 9
## # Groups:   PartNo, SectionNo, HouseNo [6]
##   PartNo SectionNo HouseNo VoterNameEn Sex   Age   FirstName MiddleName LastName
##   <chr>  <chr>     <chr>   <chr>       <chr> <chr> <chr>     <chr>      <chr>
## 1 1      2         28      Pramod Sin~ M     45    Pramod    <NA>       Singh
## 2 1      2         29      Amaresh Sih M     64    Amaresh   <NA>       Sih
## 3 1      2         30      SITARAM SI~ M     51    SITARAM   <NA>       Singh
## 4 1      2         31      Avadhesh S~ M     53    Avadhesh  <NA>       Sinh
## 5 1      2         32      Chandeshwa~ M     50    Chandesh~ <NA>       Mehato
## 6 1      2         33      Ganga Pras~ M     72    Ganga     Prasad     Chauras~
```

```
df2_badh <- df2_badh %>% group_by(LastName) %>% mutate(count=n())
surname_badh <- df2_badh %>% select(LastName,count)
surname_badh <- unique(surname_badh)
surname_badh<- surname_badh %>% arrange(desc(count))
head(surname_badh)
```

```
## # A tibble: 6 x 2
## # Groups:   LastName [6]
##   LastName count
##   <chr>    <int>
## 1 Singh     5184
## 2 Paswan    3081
## 3 Prasad    2898
## 4 Yadav     1905
## 5 Mahto     1810
## 6 Ray       1667
```

```
top_surname_badh <- head(surname_badh)

label <- top_surname_badh$LastName[1:6]
per <- top_surname_badh$count[1:6]/sum(top_surname_badh$count[1:6])*100
per <- as.integer(per)
pie(top_surname_badh$count[1:6],labels =label ,radius=1,main="Top 6 surnames of BADH",col=rai
nbow(6))
legend("topright",inset=0.001,legend=per,cex=0.8, fill = rainbow(6))
```


Top 6 surnames of BADH
Singh
Paswan
Ray
Prasad
Mahto
Yadav
31
18
17
11
10
10