Assignment 1

By Riya Gupta (17BCS026)

The dataset under investigation is the Live Births by Occupation of the Mother and Birth Order (Rural & Urban) for years 2011 to 2016

There are 6 tables, one for each year, all comprising of 17 columns in total. The first column is for Occupation of the Mother, column 2 to 14 for birth order, 15th for not stated, the 16th is for total for an occupation (in all areas or in rural/urban areas) and the last is for the area.

Birth order refers to the order a child is born in their family; first-born and second-born are examples. Birth order is often believed to have a profound and lasting effect on psychological development.

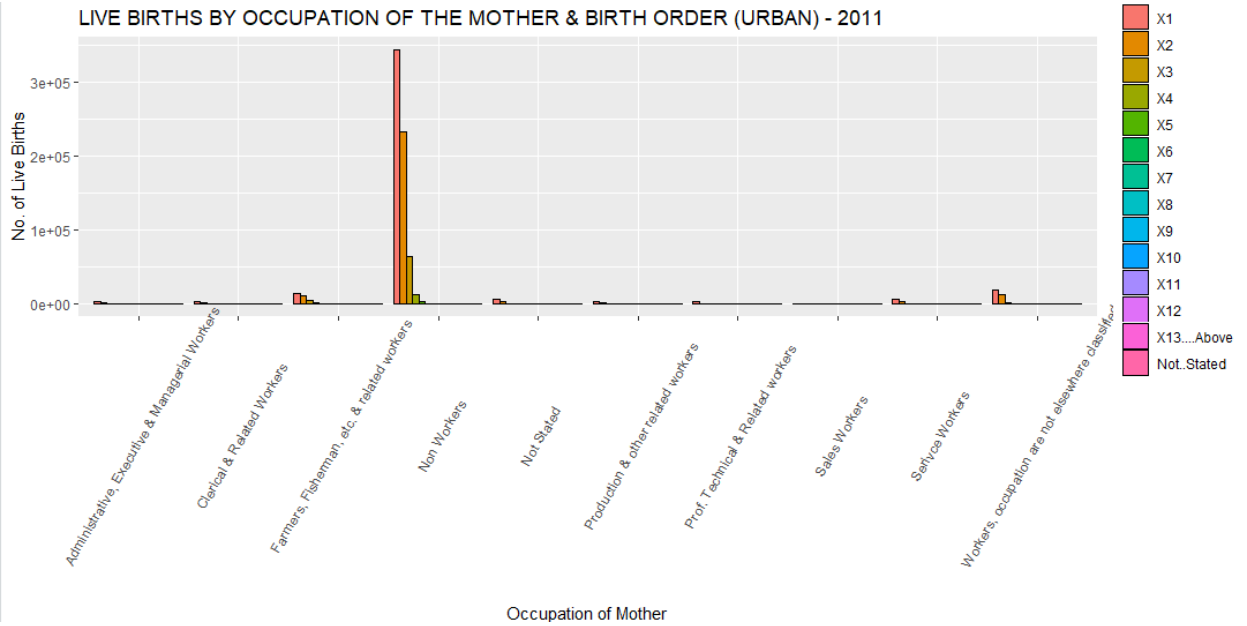
To extract the required data, I opened the PDF in Word by dragging and dropping in the new window of Word. In Word I looked for the required table and copied it to paste in an Excel sheet where I modified the little extras and then saved this document in CSV format.

Following is the screenshot of 2013 table in CSV format.

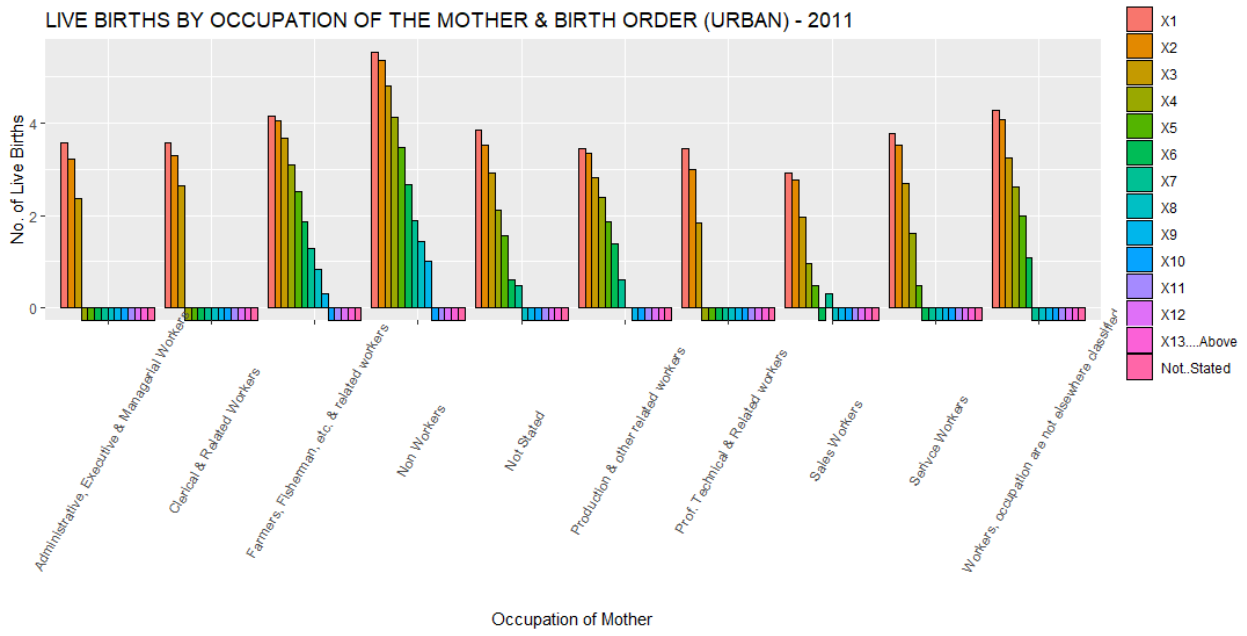
All Areas	1	2	3	4	5	6	7	8	9	10	11	12	13 &	Not	Total	Area
Prof. Technical	12828	5806	55	0	0	0	0	0	0	0	0	0	0	0	18689	All
Administrative,	8079	4496	64	0	0	0	0	0	0	0	0	0	0	0	12639	All
Clerical & Relat	6220	3629	92	0	0	0	0	0	0	0	0	0	0	0	9941	All
Sales Workers	5230	3540	183	14	4	0	0	0	0	0	0	0	0	0	8971	All
Serivce Worker	33016	21682	8543	1529	662	141	51	18	20	0	0	0	0	0	65662	All
Farmers, Fisher	65046	42645	15431	2438	1002	224	131	45	43	0	0	0	0	0	127005	All
Production & o	2576	1983	442	15	3	1	1	0	0	0	0	0	0	0	5021	All
Workers, occup	20526	15357	5403	655	262	46	33	10	10	0	0	0	0	0	42302	All
Non Workers	403347	284690	73107	4253	1603	239	106	62	72	0	0	0	0	0	767479	All
Not Stated	4971	3726	1861	281	91	9	10	8	5	0	0	0	0	0	10962	All
TOTAL	561839	387554	105181	9185	3627	660	332	143	150	0	0	0	0	0	1068671	All
Prof. Technical	12232	5426	43	0	0	0	0	0	0	0	0	0	0	0	17701	Urban
Administrative,	7158	3787	46	0	0	0	0	0	0	0	0	0	0	0	10991	Urban
Clerical & Relat	5483	3063	69	0	0	0	0	0	0	0	0	0	0	0	8615	Urban
Sales Workers	4413	2943	138	11	2	0	0	0	0	0	0	0	0	0	7507	Urban
Serivce Worker	23070	14336	4686	471	213	25	9	0	8	0	0	0	0	0	42818	Urban
Farmers, Fisher	35855	20313	5356	387	184	21	19	2	8	0	0	0	0	0	62145	Urban
Production & o	2139	1618	303	6	1	0	1	0	0	0	0	0	0	0	4068	Urban
Workers, occup	11069	7388	2047	168	70	10	7	3	3	0	0	0	0	0	20765	Urban
Non Workers	328460	222182	50880	1844	832	78	19	20	50	0	0	0	0	0	604365	Urban
Not Stated	4039	3077	1469	122	37	1	7	6	5	0	0	0	0	0	8763	Urban
TOTAL	433918	284133	65037	3009	1339	135	62	31	74	0	0	0	0	0	787738	Urban
Prof. Technical	596	380	12	0	0	0	0	0	0	0	0	0	0	0	988	Rural
Administrative,	921	709	18	0	0	0	0	0	0	0	0	0	0	0	1648	Rural
Clerical & Relat	737	566	23	0	0	0	0	0	0	0	0	0	0	0	1326	Rural
Sales Workers	817	597	45	3	2	0	0	0	0	0	0	0	0	0	1464	Rural
Serivce Worker	9946	7346	3857	1058	449	116	42	18	12	0	0	0	0	0	22844	Rural
Farmers, Fisher	29191	22332	10075	2051	818	203	112	43	35	0	0	0	0	0	64860	Rural

```
library(reshape2)

library(ggplot2)

library(dplyr)
library(stats)

data11 <- read.csv("C:/Users/Riya/Desktop/DA-lab1/CRS-2011.csv")
data12 <- read.csv("C:/Users/Riya/Desktop/DA-lab1/CRS-2012.csv")
```

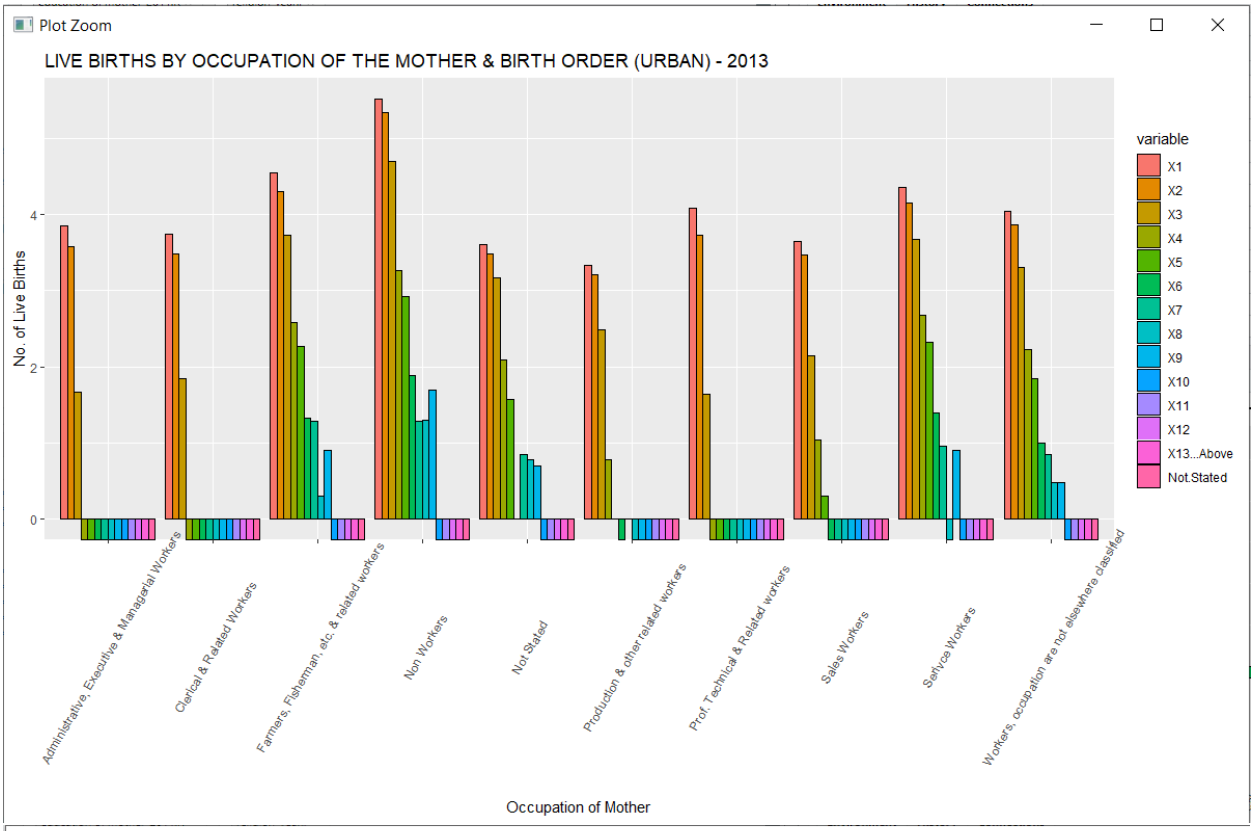
```
data13 <- read.csv("C:/Users/Riya/Desktop/DA-lab1/CRS-2013.csv")
data14 <- read.csv("C:/Users/Riya/Desktop/DA-lab1/CRS-2014.csv")
data15 <- read.csv("C:/Users/Riya/Desktop/DA-lab1/CRS-2015.csv")
data16 <- read.csv("C:/Users/Riya/Desktop/DA-lab1/CRS-2016.csv")
```

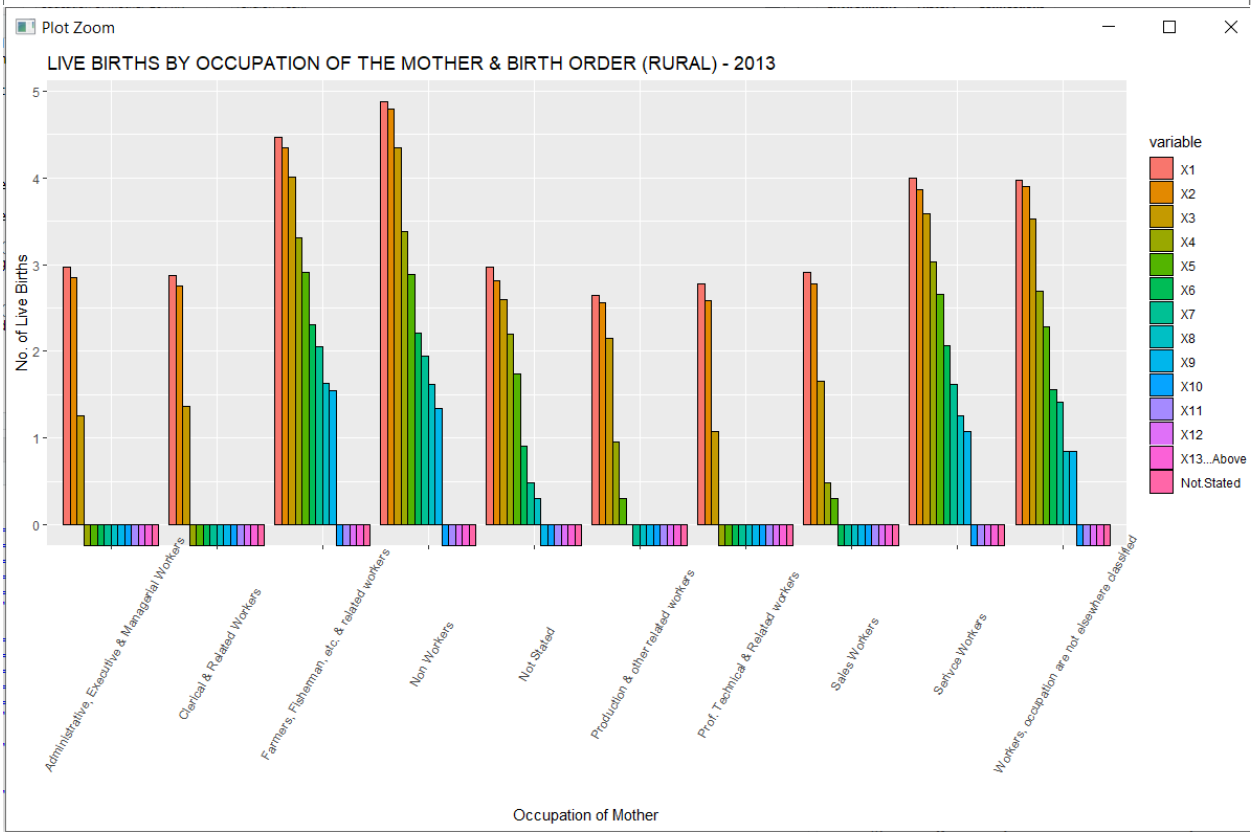


The values of number of Live Births is plotted against occupation of mother.



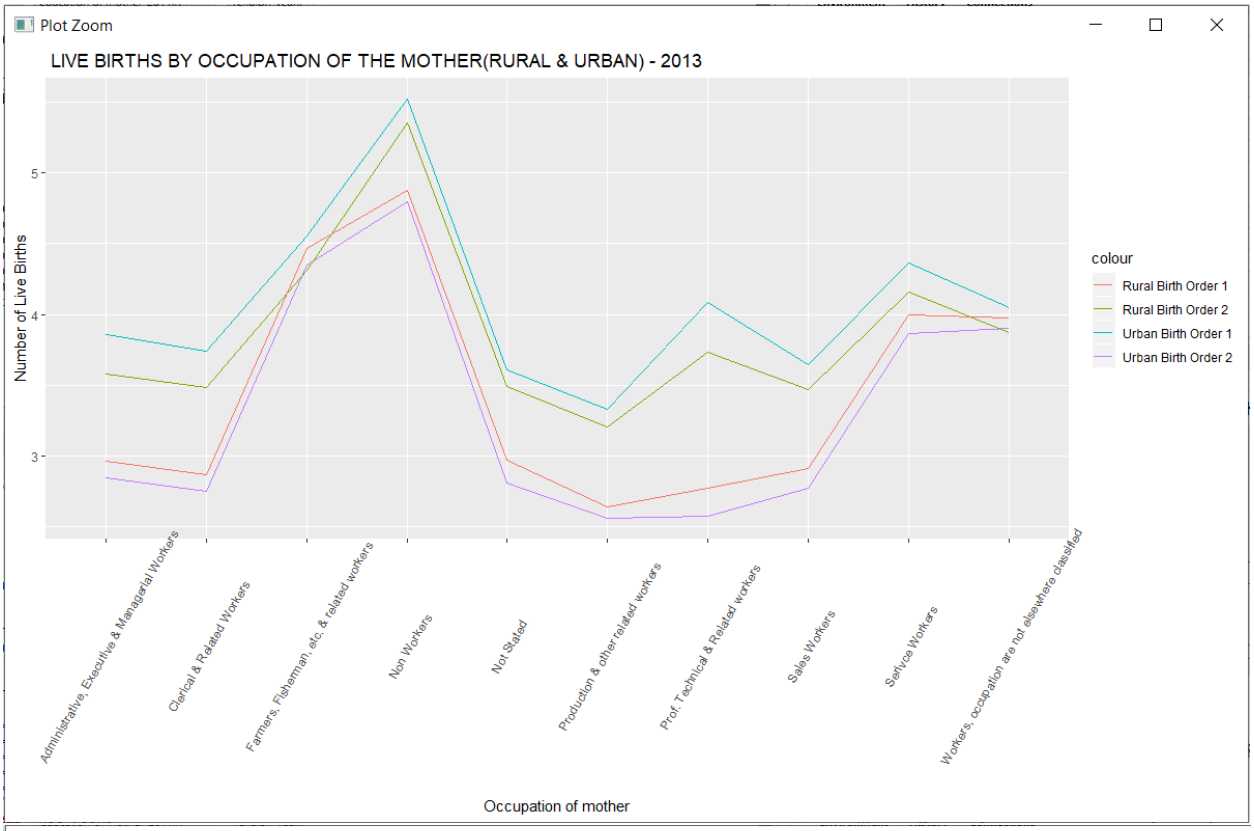
Same plot but the number of live births is normalized by taking log10.

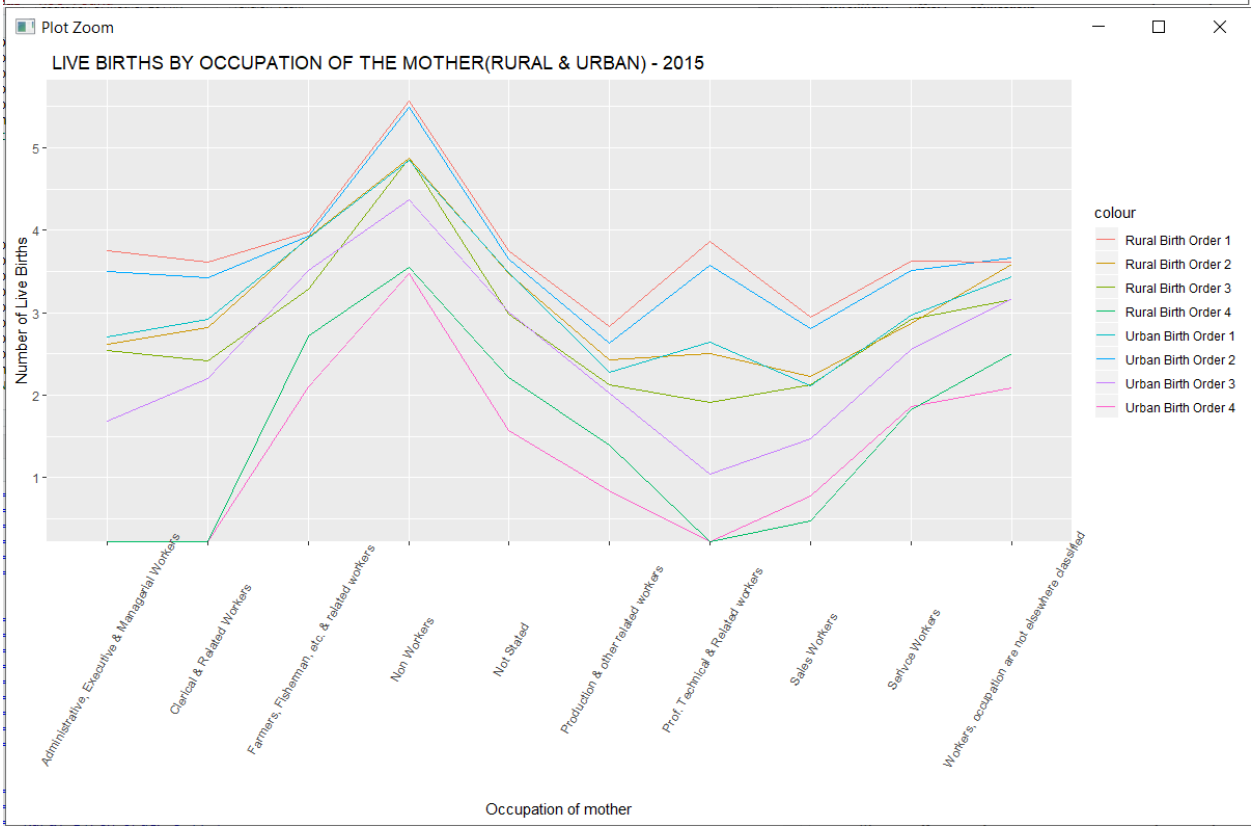

Plot Zoom
LIVE BIRTHS BY OCCUPATION OF THE MOTHER & BIRTH ORDER (URBAN) - 2013
No. of Live Births
0
2
4
variable
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12
X13...Above
Not.Stated
Administrative, Executive & Managerial Workers
Clerical & Related Workers
Farmers, Fisherman, etc. & related workers
Non Workers
Not Stated
Production & other related workers
Prof. Technical & Related workers
Sales Workers
Sehrvce Workers
Workers, occupation are not elsewhere classified
Occupation of Mother

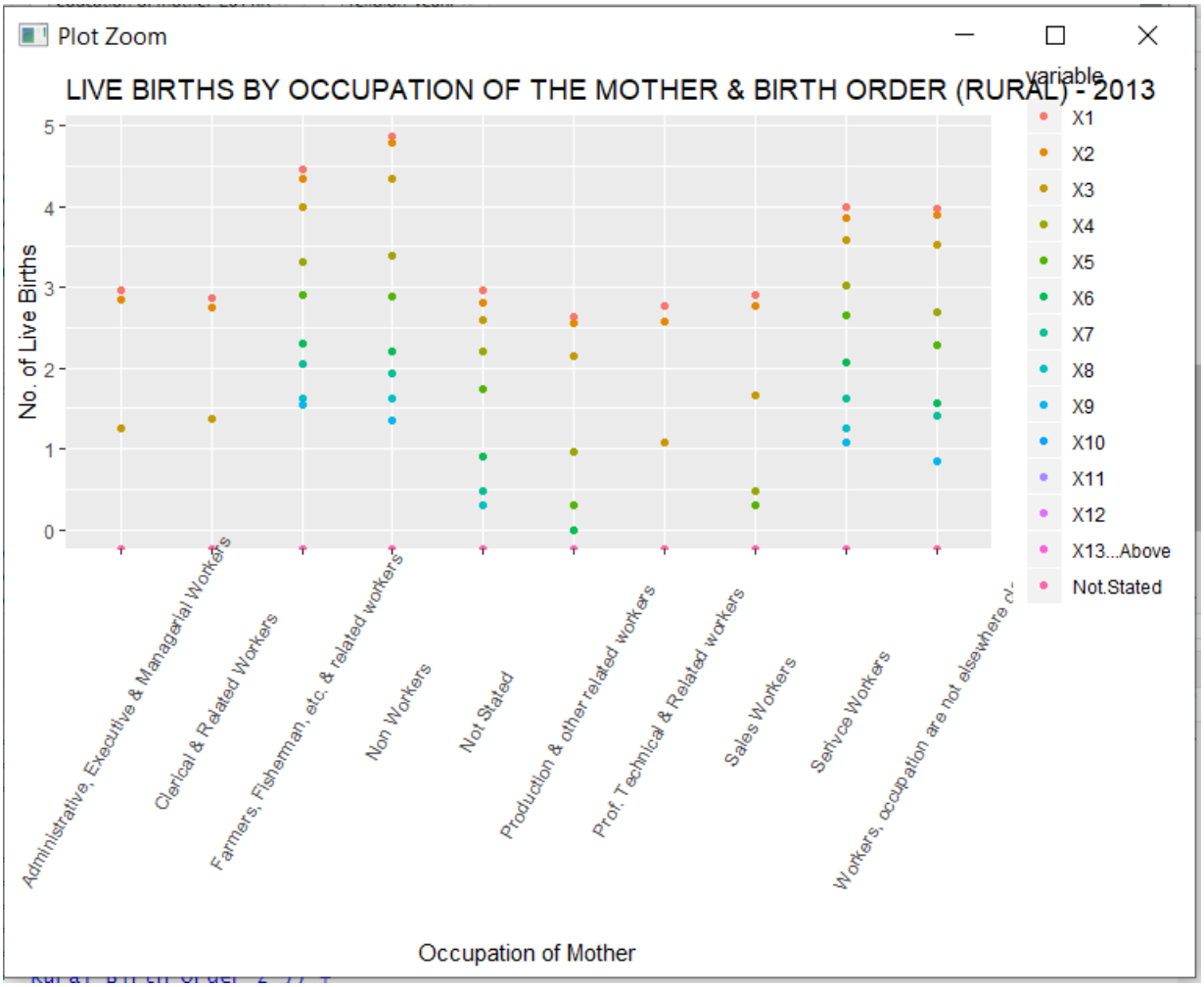

Plot Zoom
LIVE BIRTHS BY OCCUPATION OF THE MOTHER & BIRTH ORDER (RURAL) - 2013
No. of Live Births
0
1
2
3
4
5
variable
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12
X13...Above
Not.Stated
Administrative, Executive & Managerial Workers
Clerical & Related Workers
Farmers, Fisherman, etc. & related workers
Non Workers
Not Stated
Production & other related workers
Prof. Technical & Related workers
Sales Workers
Sehrvce Workers
Workers, occupation are not elsewhere classified
Occupation of Mother

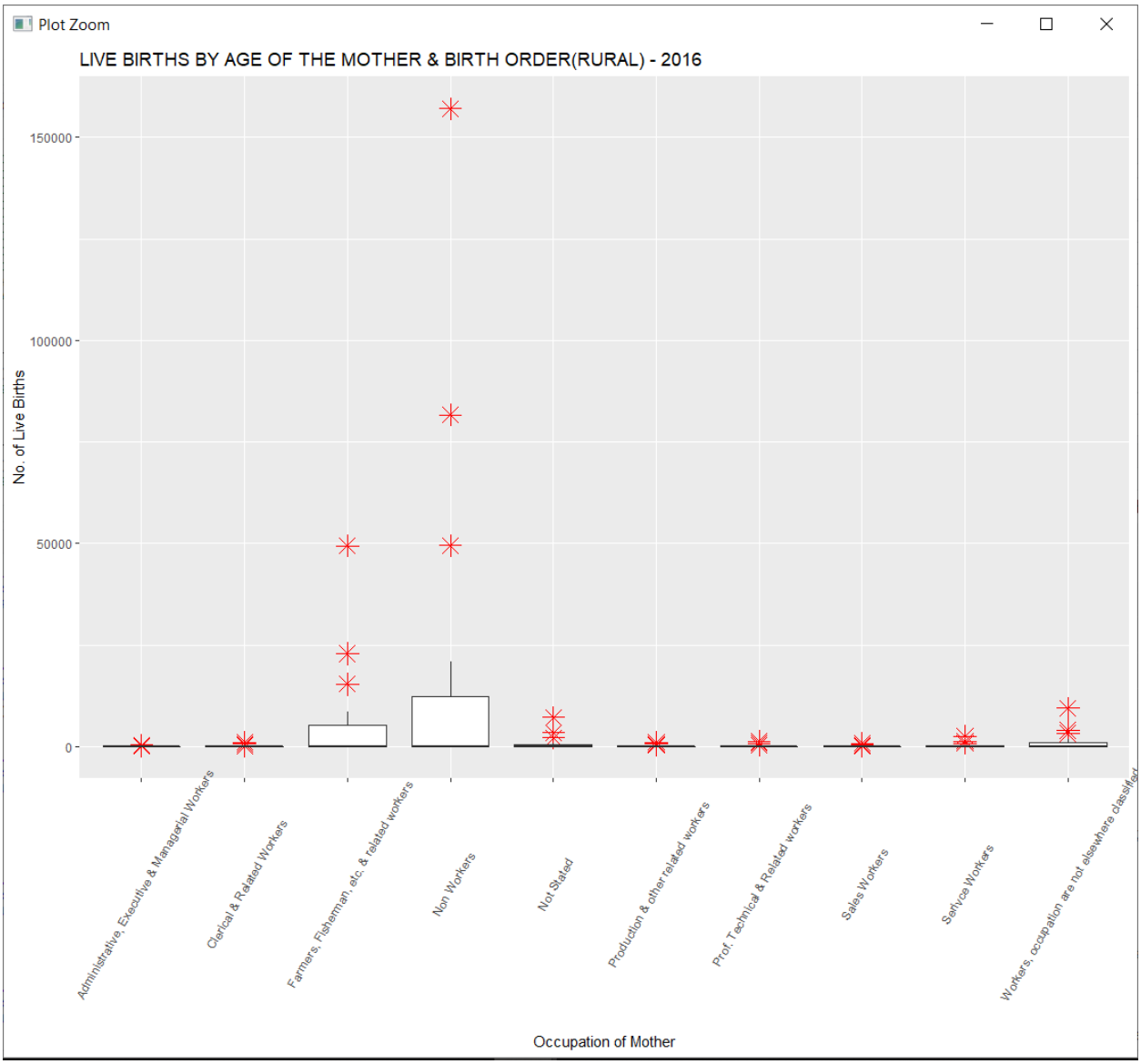
```
> summary(data15)
                                        ALL.AREAS        X1              X2              X3              X4
 Administrative, Executive & Managerial Workers: 3   Min.   :   131   Min.   :   170   Min.   :    11   Min.   :    0
 Clerical & Related Workers                     : 3   1st Qu.:   924   1st Qu.:   745   1st Qu.:   158   1st Qu.:    0
 Farmers, Fisherman, etc. & related workers     : 3   Median :  5093   Median :  3774   Median :   832   Median :  115
 Non Workers                                    : 3   Mean   : 59758   Mean   : 51770   Mean   : 13500   Mean   : 1840
 Not Stated                                     : 3   3rd Qu.:  8644   3rd Qu.:  8376   3rd Qu.:  2908   3rd Qu.:  519
 Production & other related workers             : 3   Max.   :492999   Max.   :427103   Max.   :111375   Max.   :15181
 (Other)                                        :15
       X5                X6                X7                X8                X9          X10               X11               X12
 Min.   :   0.0   Min.   :   0.0   Min.   :  0.00   Min.   :  0.00   Min.   :0   Min.   :0.0000   Min.   :0.0000   Min.   :0
 1st Qu.:   0.0   1st Qu.:   0.0   1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:0   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0
 Median :  37.0   Median :   4.0   Median :  4.00   Median :  0.00   Median :0   Median :0.0000   Median :0.0000   Median :0
 Mean   : 600.5   Mean   : 150.9   Mean   : 32.24   Mean   : 14.91   Mean   :0   Mean   :0.1212   Mean   :0.1212   Mean   :0
 3rd Qu.: 235.0   3rd Qu.:  63.0   3rd Qu.: 24.00   3rd Qu.: 10.00   3rd Qu.:0   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0
 Max.   :4954.0   Max.   :1245.0   Max.   :266.00   Max.   :123.00   Max.   :0   Max.   :1.0000   Max.   :1.0000   Max.   :0

  X13...Above    Not.Stated      Total             Area
 Min.   :0     Min.   :0     Min.   :    335   All  :11
 1st Qu.:0     1st Qu.:0     1st Qu.:   2029   Rural:11
 Median :0     Median :0     Median :  10046   Urban:11
 Mean   :0     Mean   :0     Mean   : 127666
 3rd Qu.:0     3rd Qu.:0     3rd Qu.:  20338
 Max.   :0     Max.   :0     Max.   :1053248

>
> summary(data14)
                                        All.Areas        X1              X2              X3              X4
 Administrative, Executive & Managerial Workers: 3   Min.   :   383   Min.   :   292   Min.   :    0   Min.   :    0
 Clerical & Related Workers                     : 3   1st Qu.:  2403   1st Qu.:  2025   1st Qu.:    0   1st Qu.:    0
 Farmers, Fisherman, etc. & related workers     : 3   Median :  7122   Median :  6180   Median :  1330   Median :   91
 Non Workers                                    : 3   Mean   : 66174   Mean   : 50961   Mean   : 12666   Mean   : 1500
 Not Stated                                     : 3   3rd Qu.: 21697   3rd Qu.: 14287   3rd Qu.:  5836   3rd Qu.:  639
 Production & other related workers             : 3   Max.   :545931   Max.   :420429   Max.   :104491   Max.   :12373
 (Other)                                        :15
       X5                X6                X7                X8                X9                X10               X11               X12
 Min.   :   0.0   Min.   :  0.0   Min.   :  0.00   Min.   : 0.000   Min.   : 0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0
 1st Qu.:   0.0   1st Qu.:  0.0   1st Qu.:  0.00   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0
 Median :  38.0   Median :  2.0   Median :  3.00   Median : 0.000   Median : 0.000   Median :0.0000   Median :0.0000   Median :0
 Mean   : 414.2   Mean   : 70.3   Mean   : 24.85   Mean   : 7.394   Mean   : 4.364   Mean   :0.7273   Mean   :0.1212   Mean   :0
 3rd Qu.: 169.0   3rd Qu.: 39.0   3rd Qu.: 18.00   3rd Qu.: 7.000   3rd Qu.: 6.000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0
 Max.   :3417.0   Max.   :580.0   Max.   :205.00   Max.   :61.000   Max.   :36.000   Max.   :6.0000   Max.   :1.0000   Max.   :0

  X13...Above    Not.Stated      Total             Area
 Min.   :0     Min.   :0     Min.   :    752   All  :11
 1st Qu.:0     1st Qu.:0     1st Qu.:   5206   Rural:11
 Median :0     Median :0     Median :  13303   Urban:11
 Mean   :0     Mean   :0     Mean   : 131822
 3rd Qu.:0     3rd Qu.:0     3rd Qu.:  44199
 Max.   :0     Max.   :0     Max.   :1087530
```

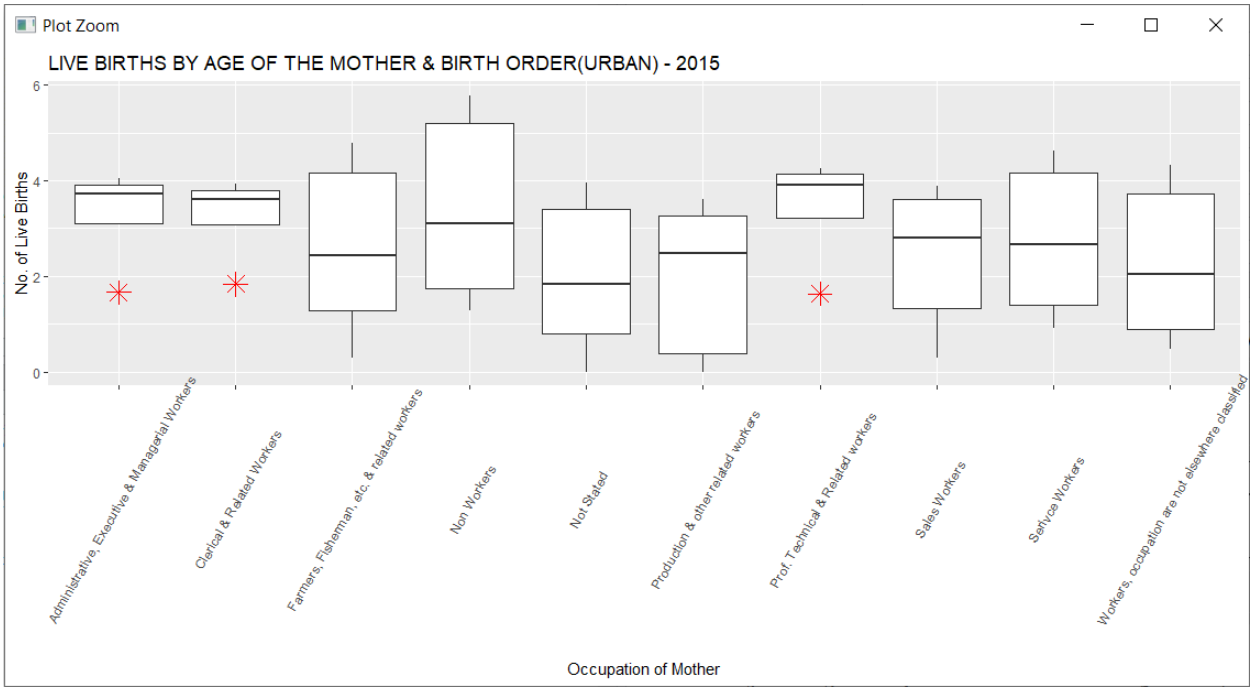
Above are the statistical analysis of years 2015 and 2014.

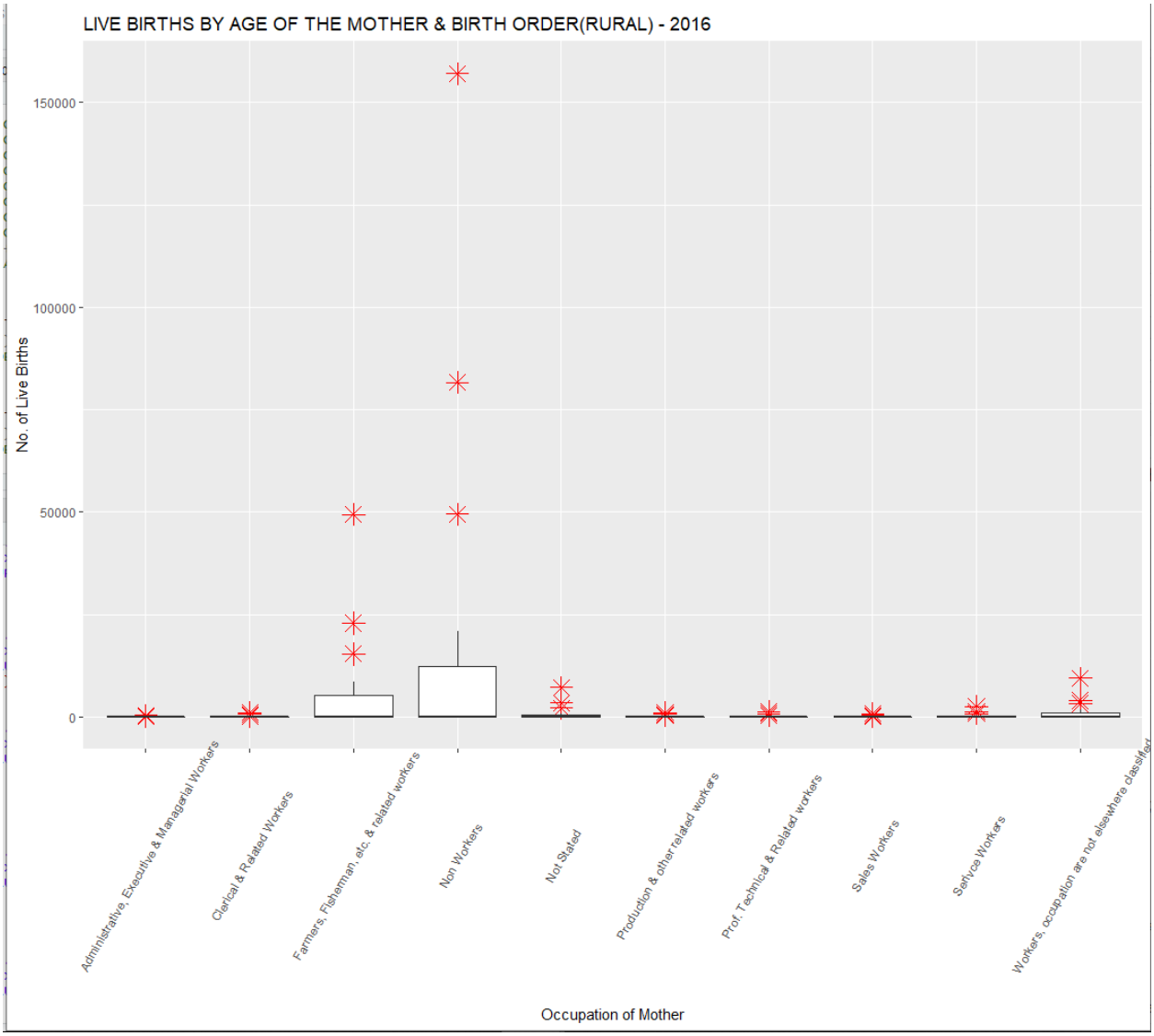

Plot Zoom
LIVE BIRTHS BY OCCUPATION OF THE MOTHER(RURAL & URBAN) - 2013
Number of Live Births
5
4
3
Administrative, Executive & Managerial Workers
Clerical & Related Workers
Farmers, Fisherman, etc. & related workers
Non Workers
Not Stated
Production & other related workers
Prof, Technical & Related workers
Sales Workers
Servics Workers
Workers, occupation are not elsewhere classified
Occupation of mother
colour
Rural Birth Order 1
Rural Birth Order 2
Urban Birth Order 1
Urban Birth Order 2


Plot Zoom
LIVE BIRTHS BY OCCUPATION OF THE MOTHER(RURAL & URBAN) - 2015
Number of Live Births
5
4
3
2
1
Administrative, Executive & Managerial Workers
Clerical & Related Workers
Farmers, Fisherman, etc. & related workers
Non Workers
Not Stated
Production & other related workers
Prof, Technical & Related workers
Sales Workers
Servics Workers
Workers, occupation are not elsewhere classified
Occupation of mother
colour
Rural Birth Order 1
Rural Birth Order 2
Rural Birth Order 3
Rural Birth Order 4
Urban Birth Order 1
Urban Birth Order 2
Urban Birth Order 3
Urban Birth Order 4


Plot Zoom
LIVE BIRTHS BY OCCUPATION OF THE MOTHER & BIRTH ORDER (RURAL) - 2013
No. of Live Births
0
1
2
3
4
5
Administrative, Executive & Managerial Workers
Clerical & Related Workers
Farmers, Fisherman, etc. & related workers
Non Workers
Not Stated
Production & other related workers
Prof. Technical & Related workers
Sales Workers
Service Workers
Occupation of Mother
variable
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12
X13...Above
Not.Stated


Plot Zoom
LIVE BIRTHS BY AGE OF THE MOTHER & BIRTH ORDER(RURAL) - 2016
150000
100000
50000
0
No. of Live Births
Administrative, Executive & Managerial Workers
Clerical & Related Workers
Farmers, Fisherman, etc. & related workers
Non Workers
Not Stated
Production & other related workers
Prof, Technical & Related workers
Sales Workers
Sehvice Workers
Occupation of Mother

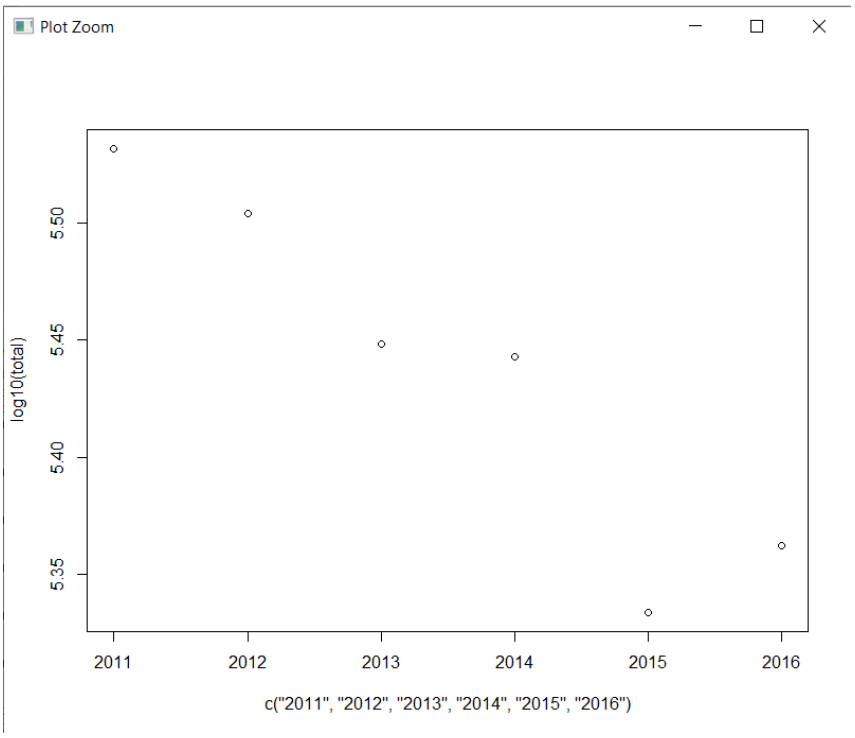

Plot Zoom
LIVE BIRTHS BY AGE OF THE MOTHER & BIRTH ORDER(URBAN) - 2015
No. of Live Births
0
2
4
6
Administrative, Executive & Managerial Workers
Clerical & Related Workers
Farmers, Fisherman, etc. & related workers
Non Workers
Not Stated
Production & other related workers
Prof, Technical & Related workers
Sales Workers
Service Workers
Workers, occupation are not elsewhere classified
Occupation of Mother



The red stars plots the outliers.

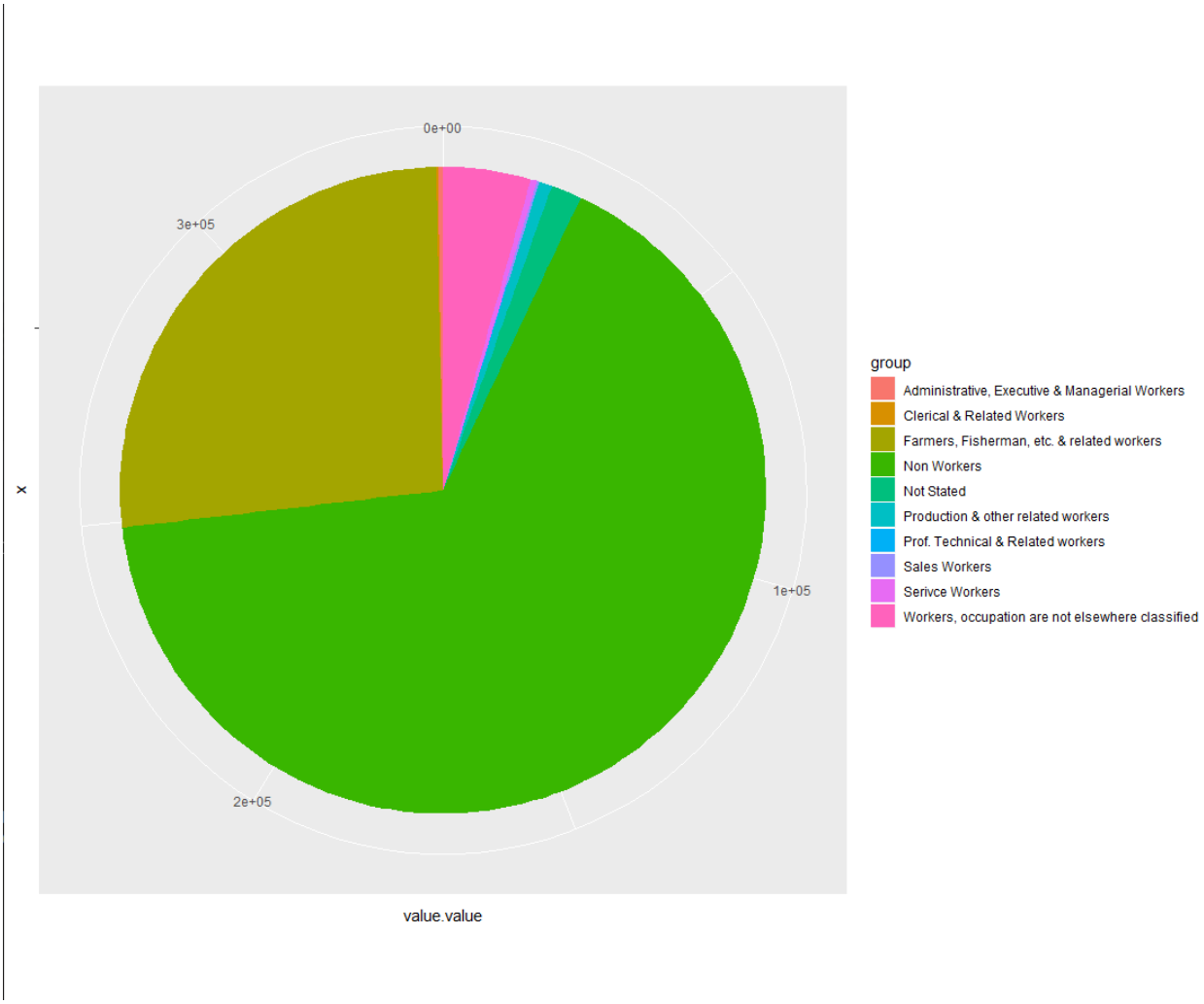
```
# total of all years - rural
w1 <- melt(data11[23:32,16])
w2 <- melt(data12[23:32,16])
w3 <- melt(data13[23:32,16])
w4 <- melt(data14[23:32,16])
w5 <- melt(data15[23:32,16])
w6 <- melt(data16[23:32,16])

ggplot(w1 , aes(x=ï..All.Areas,y=log10(value),color = variable),)+
  geom_point(stat="identity") +theme(axis.text.x = element_text(angle=60, vjust=0.6)) +
xlab("Occupation of Mother") +
```

```
  ylab("No. of Live Births") +ggtitle("LIVE BIRTHS BY OCCUPATION OF THE MOTHER & BIRTH ORDER
(RURAL) - 2013")

hh<-c("Prof. Technical & Related workers",
    + "Administrative, Executive & Managerial Workers",
    + "Clerical & Related Workers",
    + "Sales Workers",
    + "Serivce Workers",
    + "Farmers, Fisherman, etc. & related workers",
    + "Production & other related workers",
    + "Workers, occupation are not elsewhere classified",
    + "Non Workers",
    + "Not Stated")
w <- c("2011"=w1,"2012"=w2,"2013"=w3,"2014"=w4,"2015"=w5,"2016"=w6)

ww <- data.frame(w)

data <- data.frame(group=hh,value=c(w1,w2,w3,w4,w5,w6))
data

summary(data)
# rural for year 2011
mean(data$value.value)
meadian(data$value.value)
mode(data$value.value)
```



The plot is comparing total live births in rural areas over years 2011 to 2016

```
> summary(data)
                                          group    value.value       value.value.1    value.value.2    value.value.3
 Administrative, Executive & Managerial Workers:1   Min.    :   122.0   Min.    :   232   Min.    :   953   Min.    :   752
 Clerical & Related Workers                    :1   1st Qu.:   446.5   1st Qu.:   666   1st Qu.:  1360   1st Qu.:  1782
 Farmers, Fisherman, etc. & related workers    :1   Median :  1656.0   Median :  1272   Median :  1924   Median :  2794
 Non Workers                                   :1   Mean   : 34039.7   Mean   : 31927   Mean   : 28093   Mean   : 27716
 Not Stated                                    :1   3rd Qu.: 12485.5   3rd Qu.: 11347   3rd Qu.: 22517   3rd Qu.: 10112
 Production & other related workers            :1   Max.   :225183.0   Max.   :199175   Max.   :163114   Max.   :199933
 (Other)                                       :4
 value.value.4    value.value.5
 Min.    :   335   Min.    :   368
 1st Qu.:   821   1st Qu.:  1058
 Median :  1898   Median :  1905
 Mean   : 21557   Mean   : 23034
 3rd Qu.:  8260   3rd Qu.:  8920
 Max.   :172835   Max.   :157081
```




```
a1 <- melt(data11[12:21,1:15],c="All.Areas")
b1 <- melt(data11[23:32,1:15],c="All.Areas")

a <- data11 %>%
  filter(Area == "Urban" & All.Areas != "TOTAL")
b <- data11 %>%
  filter(Area == "Rural" & All.Areas != "TOTAL")

ggplot(a1,aes(x=All.Areas,y=log10(value),fill=variable),) +geom_bar(stat=
```

```
    "identity", position = position_dodge(),color = "black") +theme(axis.text.x =
    element_text(angle=60, vjust=0.6)) + xlab("Occupation of Mother") +
    ylab("No. of Live Births") +ggtitle("LIVE BIRTHS BY OCCUPATION OF THE MOTHER & BIRTH ORDER
(URBAN) - 2011")

c1 <- melt(data13[12:21,1:15],c="All.Areas")
d1 <- melt(data13[23:32,1:15],c="All.Areas")

c <- data13 %>%
  filter(Area == "Urban" & ï..All.Areas != "TOTAL")
d <- data13 %>%
  filter(Area == "Rural" & ï..All.Areas != "TOTAL")

ggplot(d1,aes(x=ï..All.Areas,y=log10(value),fill=variable),) +geom_bar(stat=
 "identity", position = position_dodge(),color = "black") +theme(axis.text.x =
  element_text(angle=60, vjust=0.6)) + xlab("Occupation of Mother") +ylab("No. of Live Births")+
  ggtitle("LIVE BIRTHS BY OCCUPATION OF THE MOTHER & BIRTH ORDER (RURAL) - 2013")
ggplot(c1,aes(x=ï..All.Areas,y=log10(value),fill=variable),) +geom_bar(stat=
  "identity", position = position_dodge(),color = "black") +theme(axis.text.x =
  element_text(angle=60, vjust=0.6)) + xlab("Occupation of Mother") +ylab("No. of Live Births")+
  ggtitle("LIVE BIRTHS BY OCCUPATION OF THE MOTHER & BIRTH ORDER (URBAN) - 2013")

summary(data14)
summary(data15)

ggplot(c, aes(x = ï..All.Areas, group = 1)) +
  geom_line(aes(y = log10(d$X2), colour = "Urban Birth Order 2")) +
  geom_line(aes(y = log10(c$X2), colour = "Rural Birth Order 2")) +
  geom_line(aes(y = log10(d$X1), colour = "Rural Birth Order 1")) +
  geom_line(aes(y = log10(c$X1), colour = "Urban Birth Order 1")) +ylab("Number of Live Births")+
  theme(axis.text.x = element_text(angle=60,
  vjust=0.6))+xlab("Occupation of mother")+ggtitle(" LIVE BIRTHS BY OCCUPATION OF THE
MOTHER(RURAL & URBAN) - 2013")

ggplot(d1 , aes(x=ï..All.Areas,y=log10(value),color = variable),)+
  geom_point(stat="identity") +theme(axis.text.x = element_text(angle=60, vjust=0.6)) +
xlab("Occupation of Mother") +
  ylab("No. of Live Births") +ggtitle("LIVE BIRTHS BY OCCUPATION OF THE MOTHER & BIRTH ORDER
(RURAL) - 2013")

e <- data15 %>%
  filter(Area == "Urban" & ALL.AREAS != "TOTAL")
f <- data15 %>%
  filter(Area == "Rural" & ALL.AREAS != "TOTAL")
```

```
ggplot(e, aes(x = ALL.AREAS, group = 1)) +
  geom_line(aes(y = log10(e$X5), colour = "Urban Birth Order 4")) +
  geom_line(aes(y = log10(f$X4), colour = "Rural Birth Order 4")) +
  geom_line(aes(y = log10(e$X3), colour = "Rural Birth Order 3")) +
  geom_line(aes(y = log10(f$X3), colour = "Urban Birth Order 3")) +
  geom_line(aes(y = log10(e$X2), colour = "Urban Birth Order 2")) +
  geom_line(aes(y = log10(f$X2), colour = "Rural Birth Order 2")) +
  geom_line(aes(y = log10(e$X1), colour = "Rural Birth Order 1")) +
  geom_line(aes(y = log10(f$X1), colour = "Urban Birth Order 1")) +ylab("Number of Live Births")+
  theme(axis.text.x = element_text(angle=60,vjust=0.6))+xlab("Occupation of mother")+
  ggtitle(" LIVE BIRTHS BY OCCUPATION OF THE MOTHER(RURAL & URBAN) - 2015")

e1 <- melt(data13[12:21,1:17],c="ï..All.Areas")
ggplot(e1 , aes(x=ï..All.Areas,y=log10(value)),) +
  geom_boxplot(outlier.colour="red",outlier.shape=8,outlier.size=5) +
  theme(axis.text.x = element_text(angle=60, vjust=0.6)) + xlab("Occupation of Mother") + ylab("No. of
Live Births") +
  ggtitle("LIVE BIRTHS BY AGE OF THE MOTHER & BIRTH ORDER(URBAN) - 2015")

h1 <- melt(data16[23:32,1:17],c="All.Areas")
ggplot(h1 , aes(x=All.Areas,y=value),) +
  geom_boxplot(outlier.colour="red",outlier.shape=8,outlier.size=5) +
  theme(axis.text.x = element_text(angle=60, vjust=0.6)) + xlab("Occupation of Mother") + ylab("No. of
Live Births") +
  ggtitle("LIVE BIRTHS BY AGE OF THE MOTHER & BIRTH ORDER(RURAL) - 2016")

ggplot(data, aes(x="", y=value.value, fill=group)) +
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)

ggplot(data, aes(x="", y=value.value.1, fill=group),
       aes(x="", y=value.value.2, fill=group)) +
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)

total <- c(sum(w1),sum(w2),sum(w3),sum(w4),sum(w5),sum(w6))
plot(c("2011","2012","2013","2014","2015","2016"),log10(total))
```