Data Analytics

Lab Assignment I

Vishal Ambavade (16CS02)

data2011.r

Vishal

Sun Feb 17 11:02:32 2019

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(ggplot2)

data2011 = read.csv("D:\\Vishal\\III year\\Data Analytics\\Assignment I\\CRS-
2011.csv",
                    nrows = 41)

data2011 <- Filter(function(x)!all(is.na(x)), data2011)

summary(data2011)

##       Year       Live_births       Still_births       Deaths
##  Min.   :1971   Min.   : 406812   Min.   : 2373   Min.   : 87556
##  1st Qu.:1981   1st Qu.: 466387   1st Qu.: 4557   1st Qu.:171857
##  Median :1991   Median : 792291   Median : 5472   Median :224115
##  Mean   :1991   Mean   : 755063   Mean   : 6227   Mean   :256804
##  3rd Qu.:2001   3rd Qu.:1009716   3rd Qu.: 6940   3rd Qu.:355662
##  Max.   :2011   Max.   :1108562   Max.   :13312   Max.   :387604
##   Vital_Births    Vital_Deaths   Percentage_Births Percentage_Deaths
```

```
##  Min.   :10.51   Min.   :2.410   Min.   :36.10     Min.   :25.10
##  1st Qu.:13.65   1st Qu.:4.590   1st Qu.:47.40     1st Qu.:45.20
##  Median :17.68   Median :6.000   Median :65.60     Median :58.40
##  Mean   :16.56   Mean   :5.572   Mean   :68.22     Mean   :66.24
##  3rd Qu.:18.95   3rd Qu.:6.640   3rd Qu.:89.34     3rd Qu.:90.71
##  Max.   :21.60   Max.   :7.160   Max.   :99.47     Max.   :98.87
```

```
head(data2011)
```

```
##   Year Live_births Still_births Deaths Vital_Births Vital_Deaths
## 1 1971      469226         9966 176160        16.00         6.00
## 2 1972      484616        13312 179593        16.10         6.00
## 3 1973      463130        11970 193725        14.94         6.24
## 4 1974      435353        10612 166102        13.74         5.24
## 5 1975      453444        10689 171857        13.97         5.29
## 6 1976      454851        10713 176061        13.68         5.30
##   Percentage_Births Percentage_Deaths
## 1              50.5              40.6
## 2              51.1              47.2
## 3              51.7              50.3
## 4              49.1              48.1
## 5              50.4              48.1
## 6              46.4              45.3
```

```
sd(data2011$Live_births)
```

```
## [1] 264177.7
```

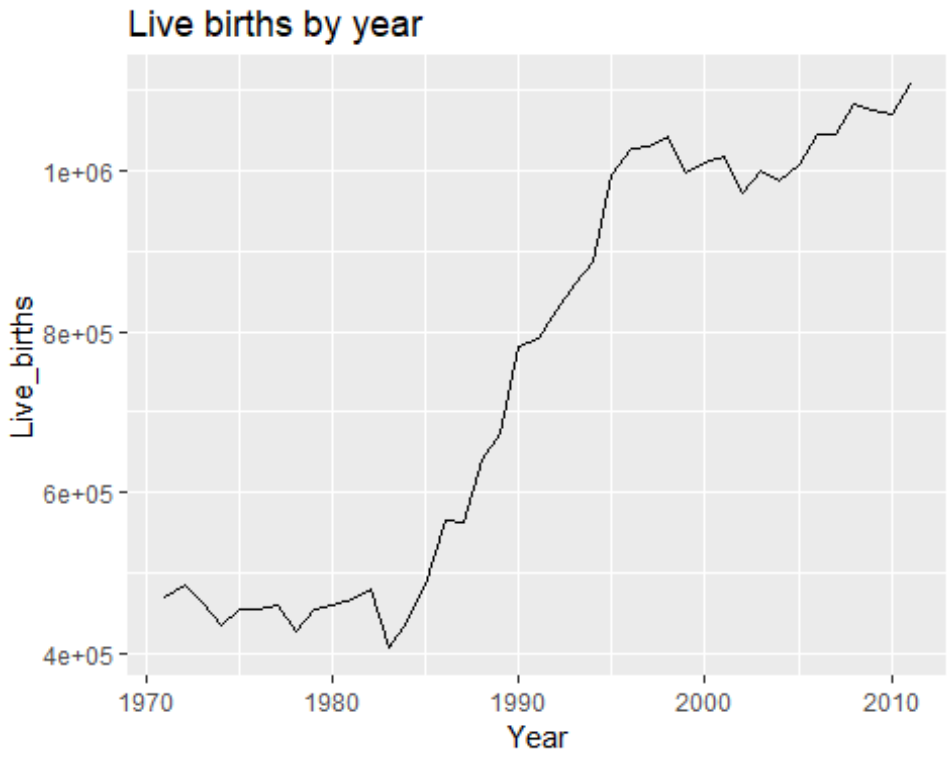
```
cor(data2011$Live_births, data2011$Deaths)
```

```
## [1] 0.9518286
```

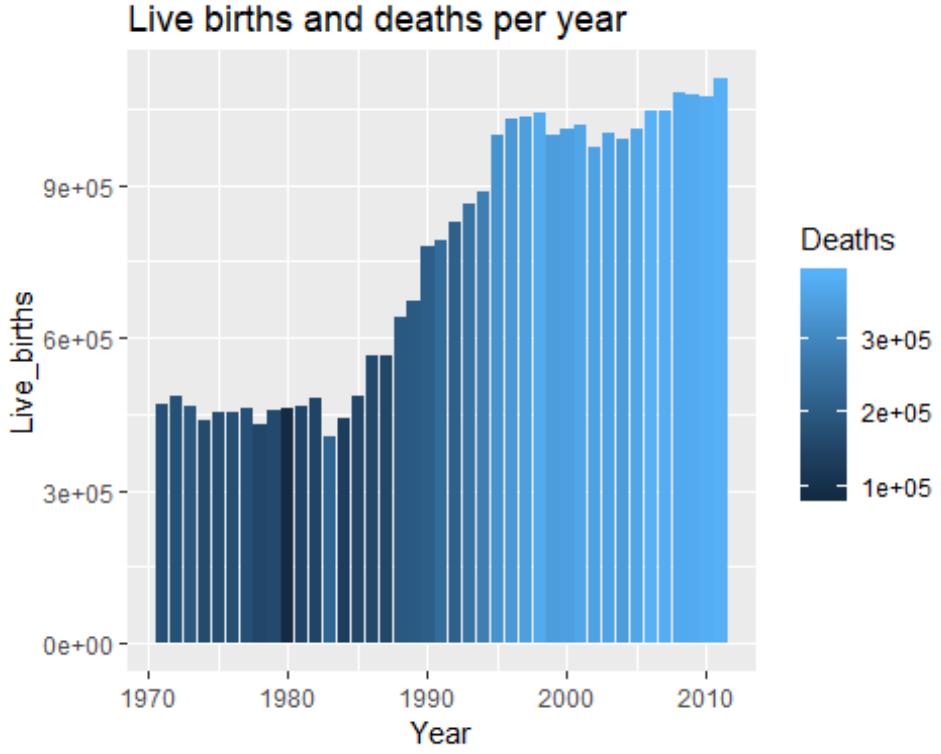
```
tail(data2011)
```

```
##    Year Live_births Still_births Deaths Vital_Births Vital_Deaths
## 36 2006     1046531         5091 387604        18.95         7.02
## 37 2007     1046424         5526 381890        18.95         6.92
## 38 2008     1082450         5069 372062        19.30         6.63
## 39 2009     1076383         5729 373290        19.05         6.61
## 40 2010     1071518         6587 381743        18.29         6.51
## 41 2011     1108562         6940 384745        18.72         6.50
##    Percentage_Births Percentage_Deaths
## 36             94.28             98.87
## 37             95.23             94.79
## 38             97.47             89.59
## 39             97.69             91.81
## 40             95.26             91.69
## 41             99.47             91.55
```

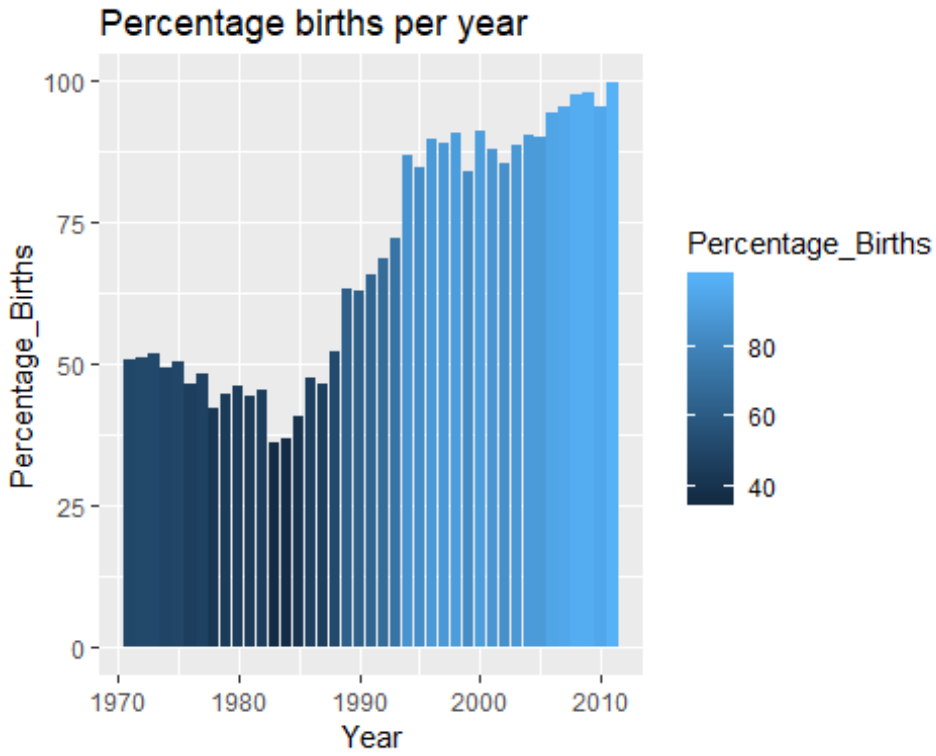
```
ggplot(data2011, aes(x = Year, y = Live_births)) +
  geom_line() +
  labs(title = "Live births by year")
```




```
ggplot(data2011, aes(fill = Deaths, x = Year, y = Live_births)) +
  geom_bar(stat = "identity") +
  labs(title = "Live births and deaths per year")
```

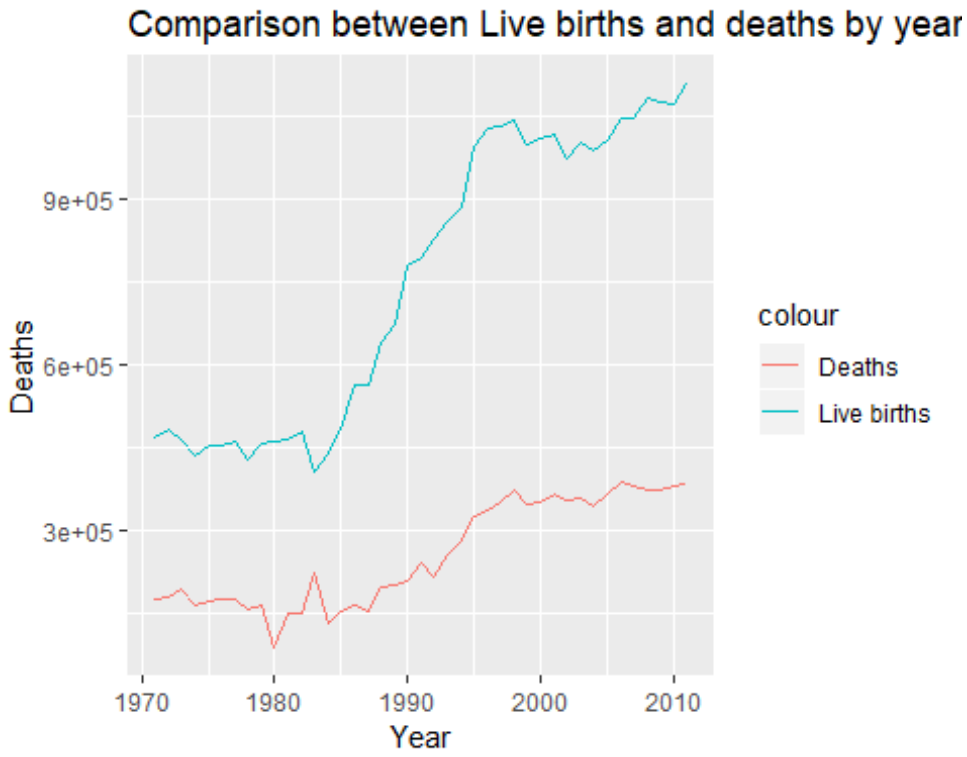


```
ggplot(data2011, aes(fill = Percentage_Births, y = Percentage_Births,
                     Percentage_Deaths, x=Year)) +
  geom_bar(position="dodge", stat="identity") +
  labs(title = "Percentage births per year")
```

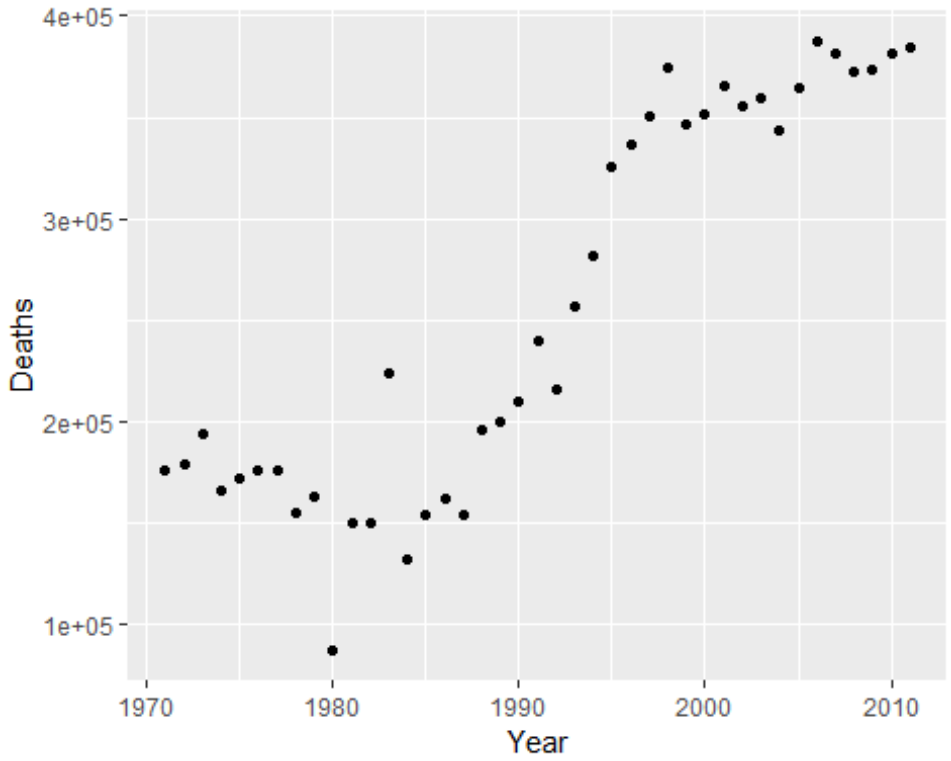



```
ggplot(data2011, aes(Year, y = Still_births, Deaths)) +
  geom_boxplot() +
  labs(title = "Still births per year")

## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```

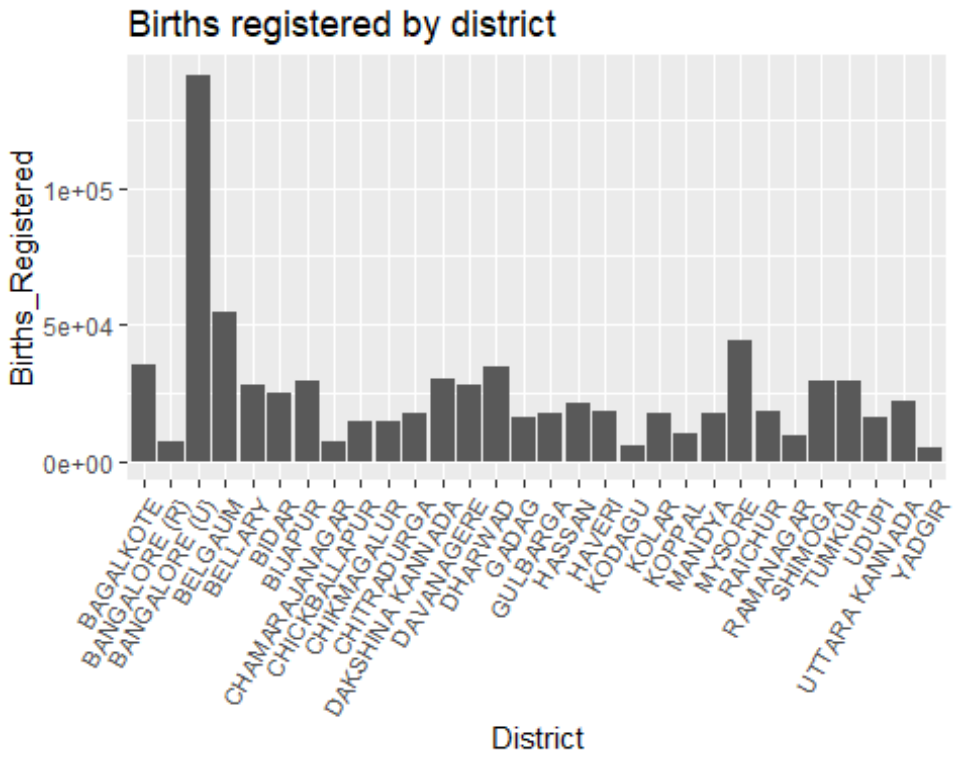

```r
ggplot(data2011, aes(Year)) +
  geom_line(aes(y = Deaths, colour = "Deaths")) +
  geom_line(aes(y = Live_births, colour = "Live births")) +
  labs(title = "Comparison between Live births and deaths by year")
```



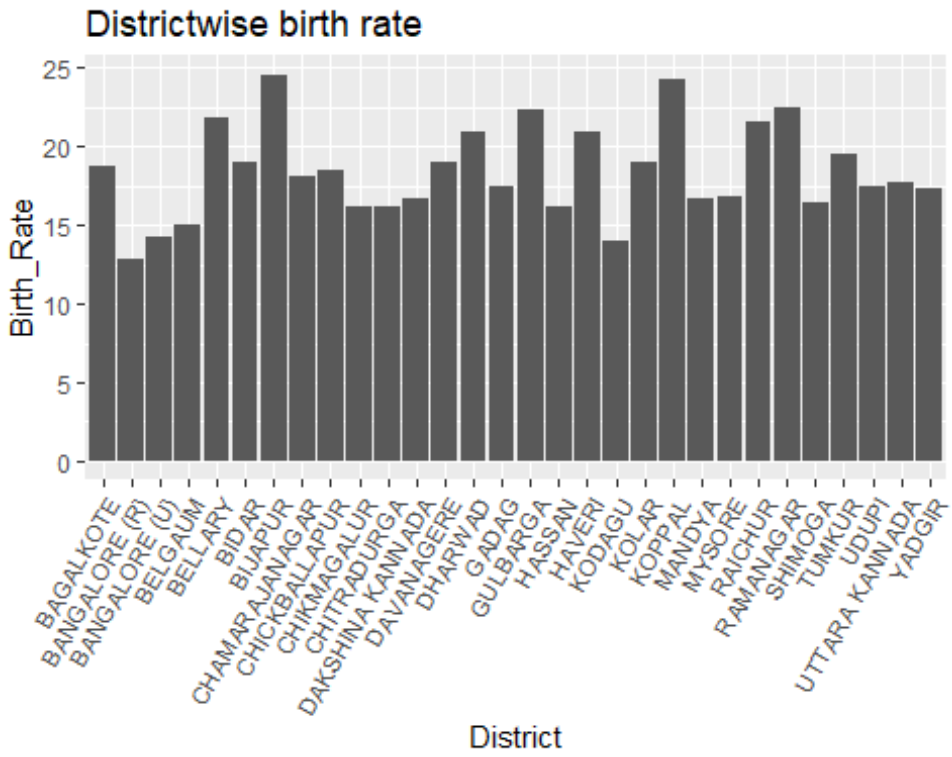

```
ggplot(data2011, aes(x = Year, y = Deaths)) +
  geom_point()
```



```r
data2011_2 <- read.csv("D:\\Vishal\\III year\\Data Analytics\\Assignment I\\CRS-2011.csv", skip=44, nrows=30,
                       header = TRUE)

data2011_2 <- Filter(function(x)!all(is.na(x)), data2011_2)
#head(data2011_2)

tb2 <- data2011_2 %>%
  filter(District != "STATE")

ggplot(tb2, aes(x = District, y = Births_Registered)) +
  geom_col() +
  theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
  labs(title = "Births registered by district")
```




```r
ggplot(tb2, aes(x = District, y = Birth_Rate)) +
  geom_col() +
  theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
  labs(title = "Districtwise birth rate")
```




```
popTable <- read.csv("D:\\Vishal\\III year\\Data Analytics\\Assignment I\\CRS
-2011.csv", skip=158, nrows=30,
                     header = TRUE)
popTable <- Filter(function(x)!all(is.na(x)), popTable)

summary(popTable)
```

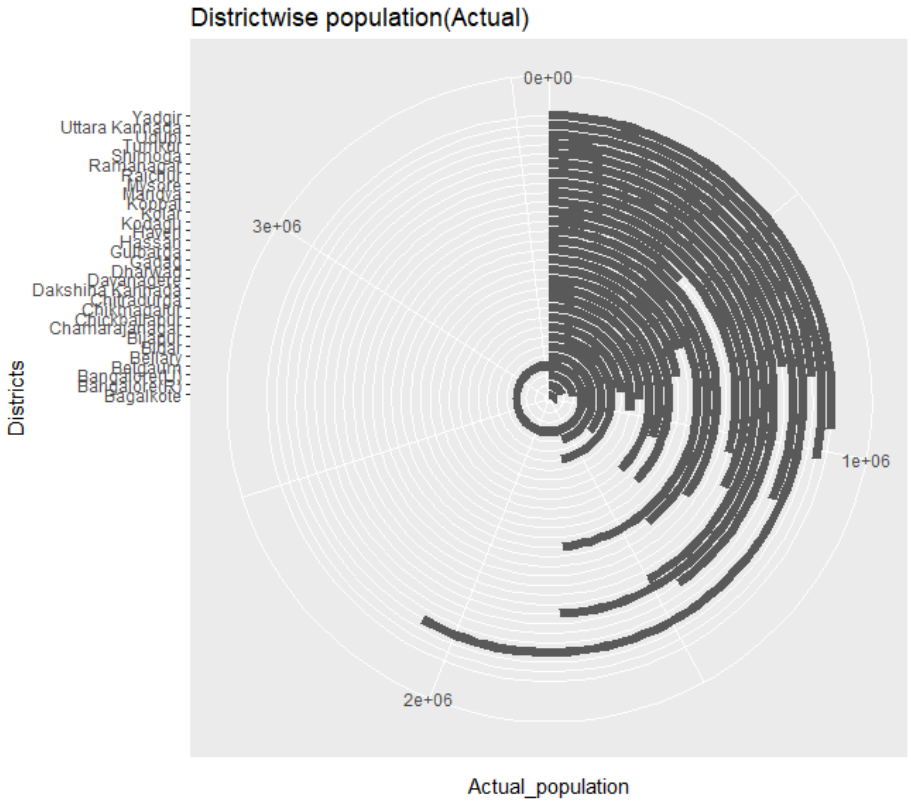
```
##      Sl.N.o             Districts  Actual_population
##  Min.   : 1.00   Bagalkote   : 1   Min.   : 473659
##  1st Qu.: 8.25   Bangalore(R): 1   1st Qu.: 876248
##  Median :15.50   Bangalore(U): 1   Median :1144972
##  Mean   :15.50   Belgaum     : 1   Mean   :1251751
##  3rd Qu.:22.75   Bellary     : 1   3rd Qu.:1427823
##  Max.   :30.00   Bidar       : 1   Max.   :3567739
##                  (Other)     :24
##  Adjusted_for_incomplete_returns No_of_Registration_Units
##  Min.   : 468192                 Min.   : 255.0
##  1st Qu.: 858540                 1st Qu.: 543.8
##  Median :1098805                 Median : 790.5
##  Mean   :1190842                 Mean   : 931.7
##  3rd Qu.:1328014                 3rd Qu.:1201.5
##  Max.   :3540968                 Max.   :2719.0
##
##  No_of_Monthly_Returns_due No_of_Monthly_Returns_not_received
##  Min.   : 3060             Min.   :   0.0
##  1st Qu.: 6525             1st Qu.:   0.0
##  Median : 9486             Median :  35.0
```

```
##  Mean   :11180             Mean   : 538.4
##  3rd Qu.:14418             3rd Qu.: 653.2
##  Max.   :32628             Max.   :3902.0
##
##  Estimated_Mid_year_Population_final Incomplete_returns
##  Min.   : 473675                     Min.   : 468207
##  1st Qu.: 878519                     1st Qu.: 859095
##  Median :1148899                     Median :1100348
##  Mean   :1254931                     Mean   :1193798
##  3rd Qu.:1432750                     3rd Qu.:1330908
##  Max.   :3580666                     Max.   :3553798
##
```

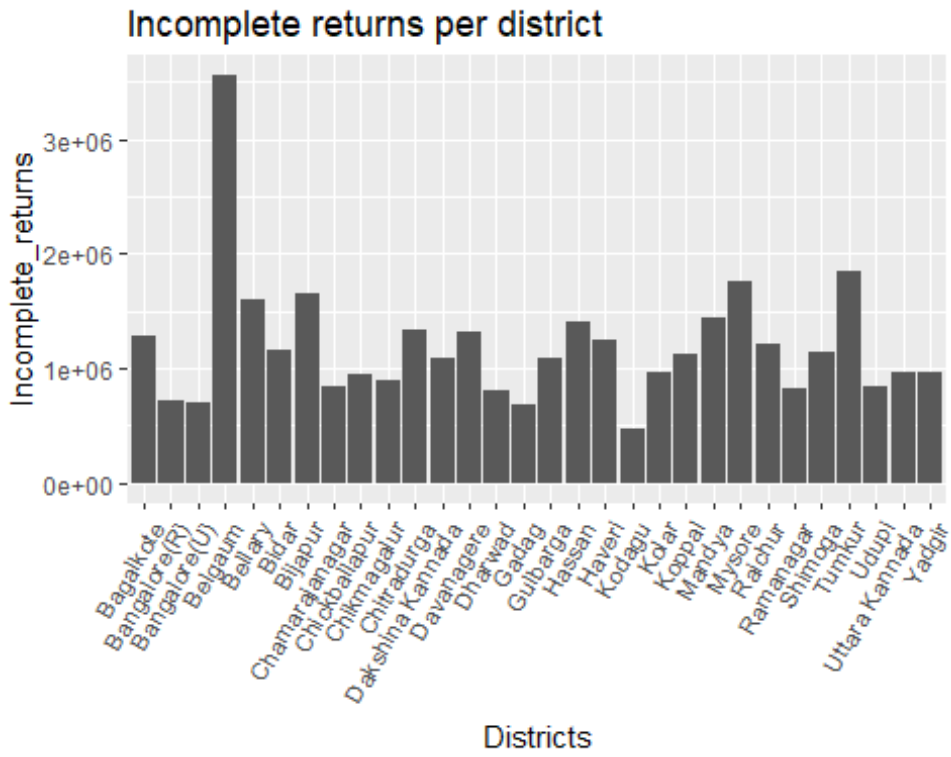
```
head(popTable)
```

```
##   Sl.N.o    Districts Actual_population Adjusted_for_incomplete_returns
## 1      1    Bagalkote           1292036                          1277161
## 2      2 Bangalore(R)            719564                           718234
## 3      3 Bangalore(U)            868971                           693938
## 4      4      Belgaum           3567739                          3540968
## 5      5      Bellary           1613038                          1586607
## 6      6        Bidar           1276647                          1161523
##   No_of_Registration_Units No_of_Monthly_Returns_due
## 1                      608                     7296
## 2                      947                    11364
## 3                      573                     6876
## 4                     1155                    13860
## 5                      534                     6408
## 6                      621                     7452
##   No_of_Monthly_Returns_not_received Estimated_Mid_year_Population_final
## 1                                 84                              1296203
## 2                                 21                               721695
## 3                               1385                               872221
## 4                                104                              3580666
## 5                                105                              1623871
## 6                                672                              1280834
##   Incomplete_returns
## 1            1281279
## 2             720361
## 3             696533
## 4            3553798
## 5            1597262
## 6            1165332
```

```
ggplot(popTable, aes(x = Districts, y = Actual_population)) +
  geom_bar(stat = "identity") +
  coord_polar(theta = "y") +
  labs(title = "Districtwise population(Actual)")
```




```
ggplot(popTable, aes(x = Districts, y = Incomplete_returns)) +
  geom_col() +
  theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
  labs(title = "Incomplete returns per district")
```




```
liveBirthsReg <- read.csv("D:\\Vishal\\III year\\Data Analytics\\Assignment I
\\CRS-2011.csv", skip=710, nrows=30,
                      header = TRUE)

liveBirthsReg <- Filter(function(x)!all(is.na(x)), liveBirthsReg)

tail(liveBirthsReg)
```

```
##     X       District RuralWithinPrescribedTimeLimitMale
## 25 25      RAMANAGAR                                1859
## 26 26        SHIMOGA                                1810
## 27 27         TUMKUR                                5493
## 28 28          UDUPI                                1698
## 29 29 UTTARA KANNADA                                1983
## 30 30         YADGIR                               11384
##    RuralWithinPrescribedTimeLimitFemale RuralWithin30Male
## 25                                 1861                10
## 26                                 1875                25
## 27                                 5447               108
## 28                                 1359                11
## 29                                 1785                23
## 30                                14936               504
##    RuralWithin30Female RuralAfter30Male RuralAfter30Female RuralAfter1Male
## 25                  20                18                 21              28
## 26                  37                30                 40             185
## 27                 115                69                 91             163
## 28                   9                22                 21             229
```

```
## 29                  10              46                 49             237
## 30                 631             163                231              82
##    RuralAfter1Female UrbanWithinPrescribedTimeLimitMale
## 25                18                               4313
## 26               120                              13794
## 27                93                              12988
## 28               116                               7738
## 29               142                              10371
## 30                30                               1130
##    UrbanWithPrescribedTimeLimitFemale UrbanWithin30Male
## 25                               4144               337
## 26                              12864               674
## 27                              12445               598
## 28                               7178               598
## 29                              10034               417
## 30                               1293               259
##    UrbanWithin30Female UrbanAfter30Male UrbanAfter30Female UrbanAfter1Male
## 25                 310              174                237             106
## 26                 631              482                529             293
## 27                 657             1213               1193             239
## 28                 558              182                133              47
## 29                 415              290                275             218
## 30                 387              605                887             111
##    UrbanAfter30Female.1
## 25                   91
## 26                  166
## 27                  163
## 28                   29
## 29                  130
## 30                   44
```

```
summary(liveBirthsReg)
```

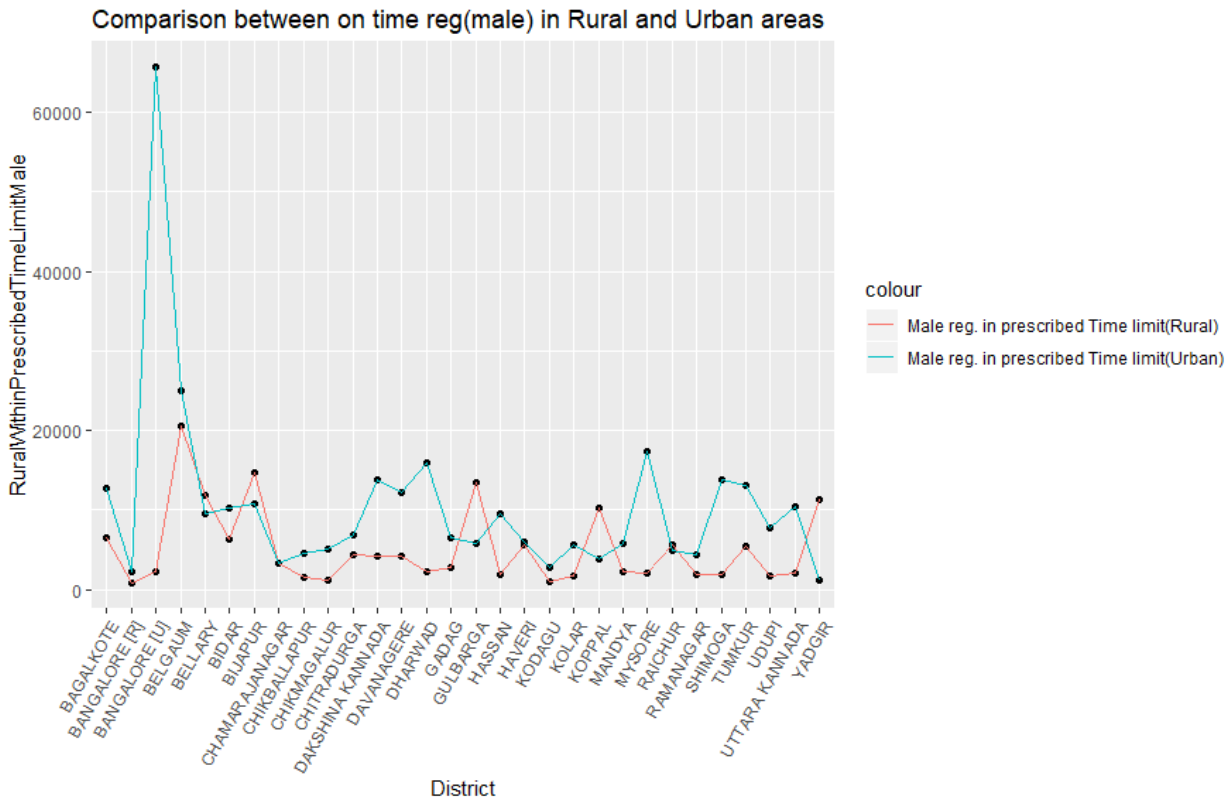
```
##        X                  District  RuralWithinPrescribedTimeLimitMale
##  Min.   : 1.00   BAGALKOTE    : 1   Min.   :  771
##  1st Qu.: 8.25   BANGALORE [R]: 1   1st Qu.: 1870
##  Median :15.50   BANGALORE [U]: 1   Median : 3058
##  Mean   :15.50   BELGAUM      : 1   Mean   : 5150
##  3rd Qu.:22.75   BELLARY      : 1   3rd Qu.: 6091
##  Max.   :30.00   BIDAR        : 1   Max.   :20531
##                  (Other)      :24
##  RuralWithinPrescribedTimeLimitFemale RuralWithin30Male
##  Min.   :  808                        Min.   : 10.0
##  1st Qu.: 1864                        1st Qu.: 32.5
##  Median : 2996                        Median :101.5
##  Mean   : 5512                        Mean   :163.1
##  3rd Qu.: 6494                        3rd Qu.:198.0
##  Max.   :19733                        Max.   :870.0
##
##  RuralWithin30Female RuralAfter30Male RuralAfter30Female RuralAfter1Male
```

```
##  Min.   :  9.00      Min.   : 18.00   Min.   : 18.00     Min.   : 17.0
##  1st Qu.: 39.75      1st Qu.: 36.25   1st Qu.: 38.00     1st Qu.: 53.5
##  Median :110.00      Median : 47.00   Median : 61.00     Median : 99.0
##  Mean   :181.57      Mean   : 69.37   Mean   : 91.97     Mean   :110.7
##  3rd Qu.:243.25      3rd Qu.: 97.25   3rd Qu.:139.75     3rd Qu.:149.8
##  Max.   :933.00      Max.   :199.00   Max.   :280.00     Max.   :305.0
##
##  RuralAfter1Female UrbanWithinPrescribedTimeLimitMale
##  Min.   : 10.00    Min.   : 1130
##  1st Qu.: 32.50    1st Qu.: 4920
##  Median : 58.00    Median : 7336
##  Mean   : 67.83    Mean   :10547
##  3rd Qu.: 92.25    3rd Qu.:12624
##  Max.   :166.00    Max.   :65676
##
##  UrbanWithPrescribedTimeLimitFemale UrbanWithin30Male UrbanWithin30Female
##  Min.   : 1293                      Min.   : 108      Min.   : 117.0
##  1st Qu.: 4593                      1st Qu.: 532      1st Qu.: 557.2
##  Median : 7015                      Median : 957      Median : 996.5
##  Mean   : 9971                      Mean   :1142      Mean   :1122.1
##  3rd Qu.:11470                      3rd Qu.:1483      3rd Qu.:1504.0
##  Max.   :61763                      Max.   :4489      Max.   :4084.0
##
##  UrbanAfter30Male UrbanAfter30Female UrbanAfter1Male  UrbanAfter30Female.1
##  Min.   :  23.0   Min.   :  21.0     Min.   :  20.0   Min.   :  12.0
##  1st Qu.: 493.2   1st Qu.: 526.8     1st Qu.:  88.5   1st Qu.:  49.0
##  Median : 972.0   Median : 890.5     Median : 128.0   Median :  89.0
##  Mean   :1175.0   Mean   :1181.6     Mean   : 275.4   Mean   : 191.6
##  3rd Qu.:1556.5   3rd Qu.:1761.2     3rd Qu.: 263.0   3rd Qu.: 141.2
##  Max.   :3731.0   Max.   :3480.0     Max.   :3601.0   Max.   :2903.0
##
```

```
cor(liveBirthsReg$RuralWithinPrescribedTimeLimitMale, liveBirthsReg$RuralWith
inPrescribedTimeLimitFemale)
```

```
## [1] 0.9864787
```

```
ggplot(liveBirthsReg, aes(x = District, group = 1)) +
  theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
  geom_point(aes(x= District, y = RuralWithinPrescribedTimeLimitMale)) +
  geom_point(aes(y = UrbanWithinPrescribedTimeLimitMale)) +
  geom_line(aes(y = RuralWithinPrescribedTimeLimitMale, colour = "Male reg. i
n prescribed Time limit(Rural)")) +
  geom_line(aes(y = UrbanWithinPrescribedTimeLimitMale, colour = "Male reg. i
n prescribed Time limit(Urban)")) +
  labs(title = "Comparison between on time reg(male) in Rural and Urban areas
")
```



data2014.r

Vishal

Sun Feb 17 11:00:33 2019

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(ggplot2)
library(reshape2)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```r
library(ggpubr)
```

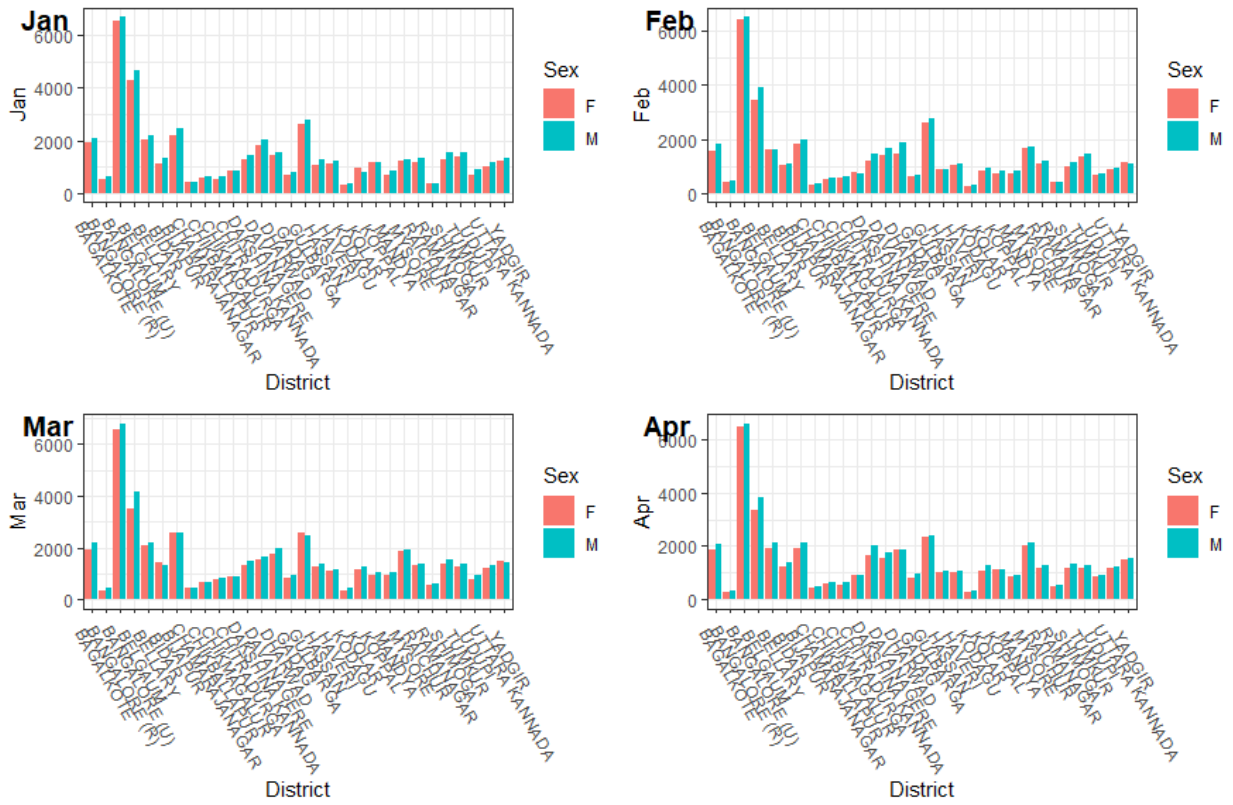
```
## Loading required package: magrittr
```

```r
data14 <- read.csv("D:\\Vishal\\III year\\Data Analytics\\Assignment I\\data14.csv", nrows = 60)


tail(data14)
```

```
##          District Sex  Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct
## 55          UDUPI   M  885  737  974  938  912  792  840  775  921  926
## 56          UDUPI   F  678  678  791  858  820  708  725  645  743  840
## 57 UTTARA KANNADA   M 1172  947 1333 1245 1146 1126 1105 1086 1149 1209
## 58 UTTARA KANNADA   F 1028  875 1209 1205 1006 1073 1106  936 1022 1096
## 59         YADGIR   M 1331 1106 1419 1525 1390 1386 1293 1303 1369 1408
## 60         YADGIR   F 1204 1123 1477 1475 1409 1446 1335 1313 1264 1424
##     Nov  Dec  X X.1
## 55  885  897 NA  NA
## 56  748  842 NA  NA
## 57  984 1303 NA  NA
## 58  907 1320 NA  NA
## 59 1215 1256 NA  NA
## 60 1259 1230 NA  NA
```

```r
data14 <- Filter(function(x)!all(is.na(x)), data14)

#df1 <- melt(data14, colnames(data14))

#df1

p1 <- ggplot(data14, aes(x = District, y = Jan)) +
  geom_bar(aes(fill = Sex), stat="identity", position ="dodge") +
  theme_bw()+
  theme(axis.text.x = element_text(angle=-60, hjust=.1))

p2 <- ggplot(data14, aes(x = District, y = Feb)) +
  geom_bar(aes(fill = Sex), stat="identity", position ="dodge") +
  theme_bw()+
  theme(axis.text.x = element_text(angle=-60, hjust=.1))
```

```
p3 <- ggplot(data14, aes(x = District, y = Mar)) +
  geom_bar(aes(fill = Sex), stat="identity", position ="dodge") +
  theme_bw()+
  theme(axis.text.x = element_text(angle=-60, hjust=.1))

p4 <- ggplot(data14, aes(x = District, y = Apr)) +
  geom_bar(aes(fill = Sex), stat="identity", position ="dodge") +
  theme_bw()+
  theme(axis.text.x = element_text(angle=-60, hjust=.1))

#ggplot(data14, aes(x=as.numeric(month), y=value, color=variable)) + geom_lin
e()

#grid.arrange(p1, p2, p3, p4)
ggarrange(p1, p2, p3, p4,
          labels = c("Jan", "Feb", "Mar", "Apr"),
          ncol = 2, nrow = 2)
```




```
data14_2 <- read.csv("D:\\Vishal\\III year\\Data Analytics\\Assignment I\\dat
a14.csv", skip = 65, header = TRUE, nrows = 27)
tail(data14_2)
```

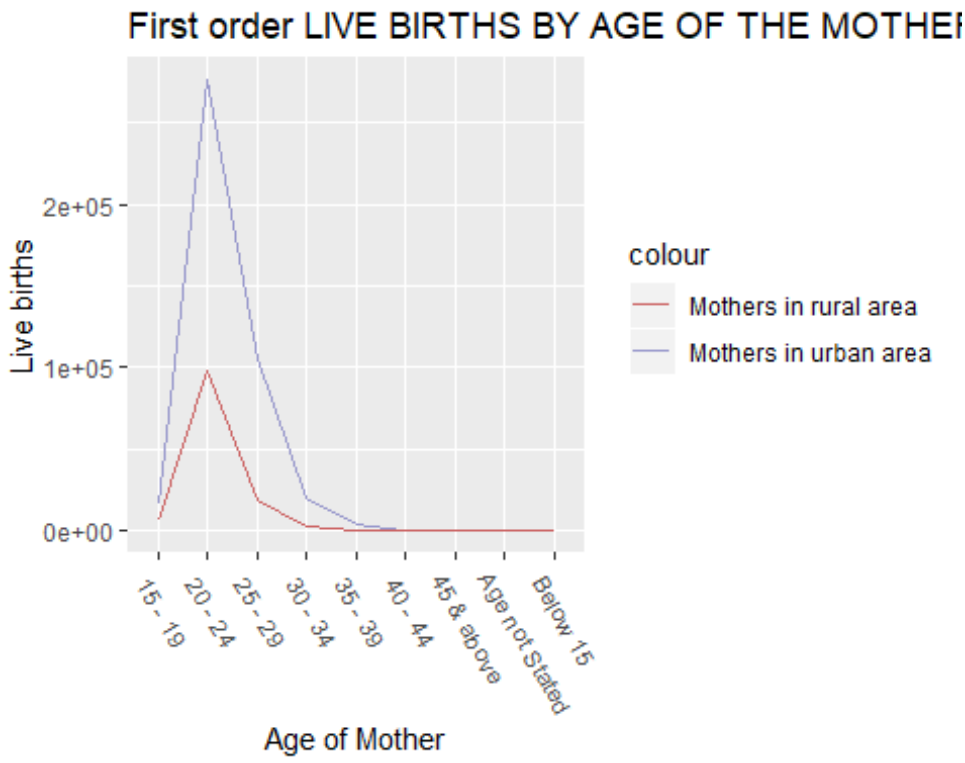
```
##     Age.of.Mother    X1    X2    X3   X4  X5  X6 X7 X8 X9 X10 X11 X12
## 22        25 - 29 18003 36395 29812 4318 721   0  0  0  0   0   0   0
## 23        30 - 34  2062  4516  3507 1385 713 193 44  0  0   0   0   0
```

```
## 24        35 - 39   432   787   636  291 208 106 58 19 16   0   0   0
## 25        40 - 44    54    69    40   27  18  19 11 13 10   1   0   0
## 26     45 & above     8     8    11   18  10   2  3  2  2   2   0   0
## 27 Age not Stated     0     0     0    0   0   0  0  0  0   0   0   0
##    X13...Above Not.Stated  Area
## 22           0          0 Rural
## 23           0          0 Rural
## 24           0          0 Rural
## 25           0          0 Rural
## 26           0          0 Rural
## 27           0          0 Rural
```

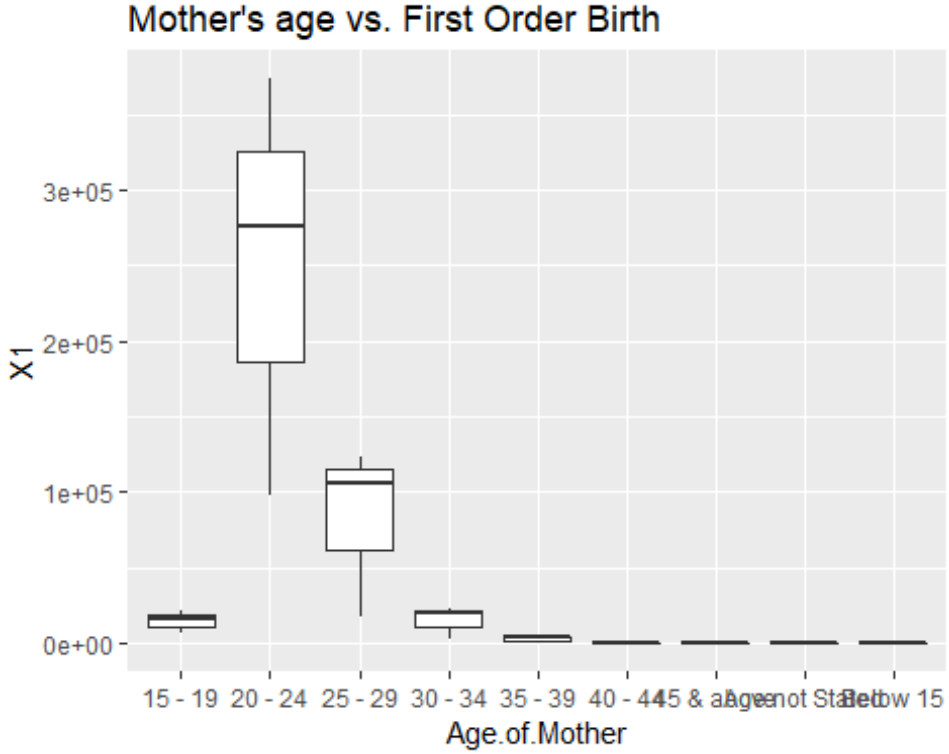
```
data14_2_urban <- data14_2 %>%
  filter(Area == "Urban")

data14_2_rural <- data14_2 %>%
  filter(Area == "Rural")

data14_2_urban$Age.of.Mother
```

```
## [1] Below 15       15 - 19        20 - 24        25 - 29
## [5] 30 - 34        35 - 39        40 - 44        45 & above
## [9] Age not Stated
## 9 Levels: 15 - 19 20 - 24 25 - 29 30 - 34 35 - 39 40 - 44 ... Below 15
```

```
ggplot(data14_2_urban, aes(x = data14_2_urban$Age.of.Mother, group = 1)) +
  geom_line(aes(y = data14_2_urban$X1, colour = "Mothers in urban area")) +
  geom_line(aes(y = data14_2_rural$X1, colour = "Mothers in rural area")) +
  theme(axis.text.x = element_text(angle=-60, hjust=.1)) +
  scale_color_manual(values=c("#CC6666", "#9999CC")) +
  ylab("Live births") + xlab("Age of Mother") +
  labs(title = "First order LIVE BIRTHS BY AGE OF THE MOTHER & BIRTH ORDER  (
RURAL & URBAN) - 2014")
```



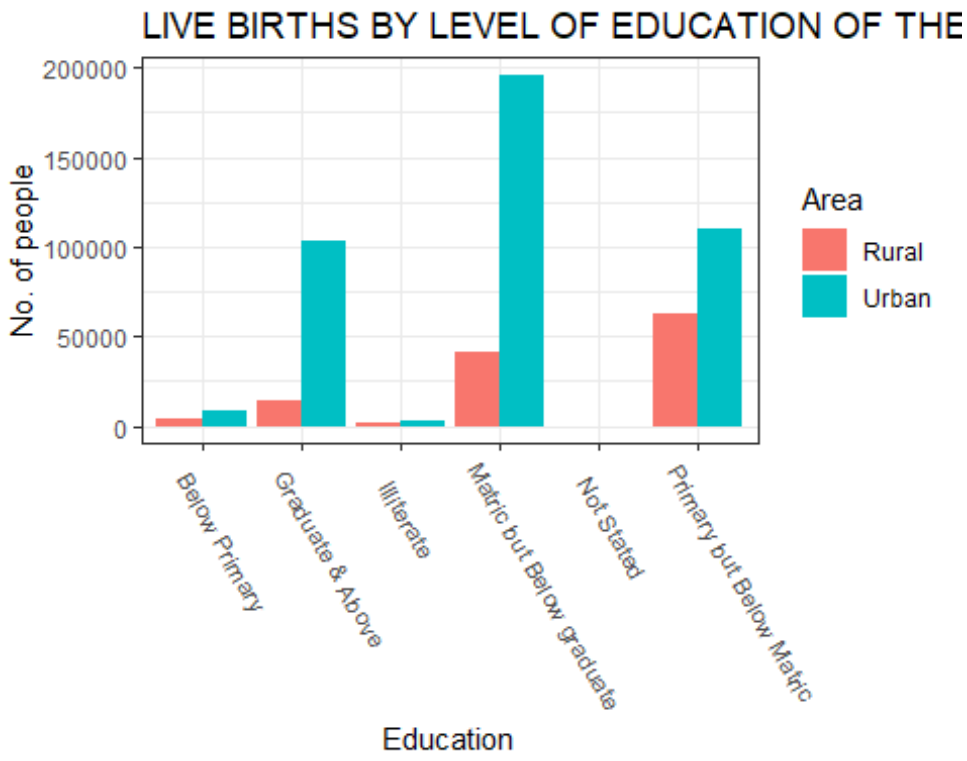

```
ggplot(data14_2, aes(x = Age.of.Mother, y = X1)) +
  geom_boxplot()  +
  labs(title = "Mother's age vs. First Order Birth")
```



```r
data14_3 <- read.csv("D:\\Vishal\\III year\\Data Analytics\\Assignment I\\dat
a14.csv", skip = 95, header = TRUE, nrows = 18)

head(data14_3)
```

```
##   Level.of.Education.of.Father    BO1    BO2   BO3  BO4  BO5 BO6 BO7 BO8
## 1                   Illiterate   5501   2083 10647 4520 1361 286 133  30
## 2                Below Primary  13739  14839  7634 3667  629 153  49  16
## 3     Primary but Below Matric 172400 140671 40999 3511  874  59   0   0
## 4    Matric but Below graduate 237216 185546 39597  675  277   0   0   0
## 5              Graduate & Above 117010  77234  5555    0    0   0   0   0
## 6                    Not Stated     65     56    59    0  276  82  23  15
##   BO9 BO10 BO11 BO12 BO13...Above Not.Stated Area
## 1  18    6    1    0            0          0  All
## 2   7    0    0    0            0          0  All
## 3   0    0    0    0            0          0  All
## 4   0    0    0    0            0          0  All
## 5   0    0    0    0            0          0  All
## 6  11    0    0    0            0          0  All
```

```r
data14_3_filter <- data14_3 %>%
  filter(Area != "All")

p31 <- ggplot(data14_3_filter, aes(x = Level.of.Education.of.Father, y = BO1)
) +
  geom_bar(aes(fill = Area), stat="identity", position ="dodge") +
  theme_bw()+
  theme(axis.text.x = element_text(angle=-60, hjust=.1)) +
  xlab("Education") + ylab("No. of people") +
  labs(title = "LIVE BIRTHS BY LEVEL OF EDUCATION OF THE FATHER & BIRTH ORDER
")

p31
```



data15.r

Vishal

Sun Feb 17 11:09:16 2019

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

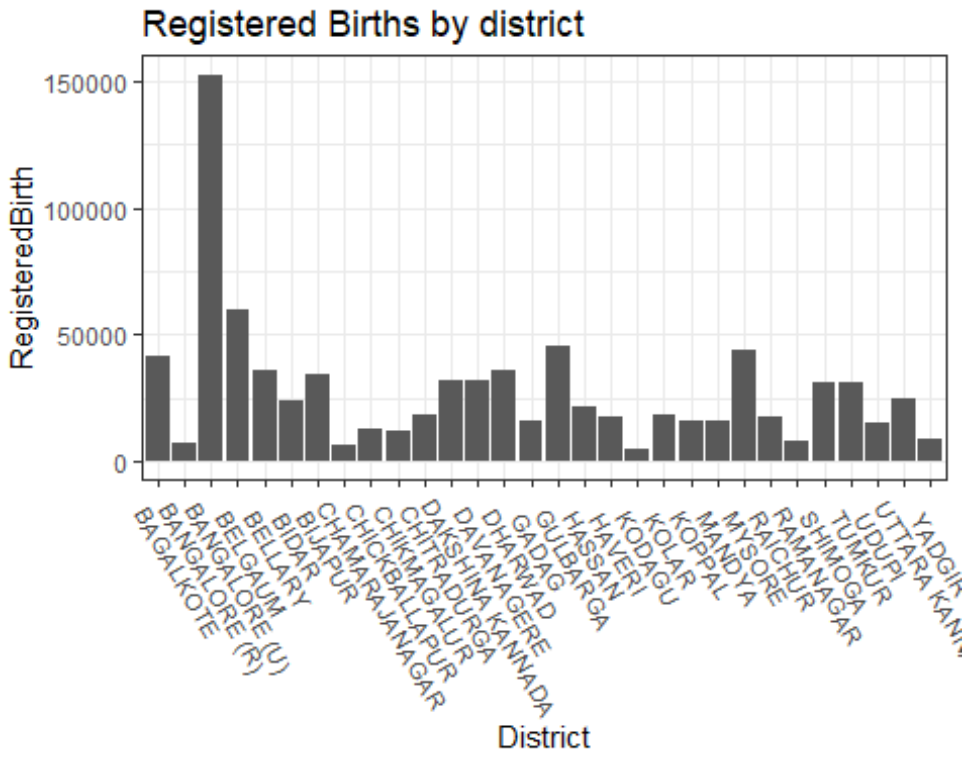
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(reshape2)
```

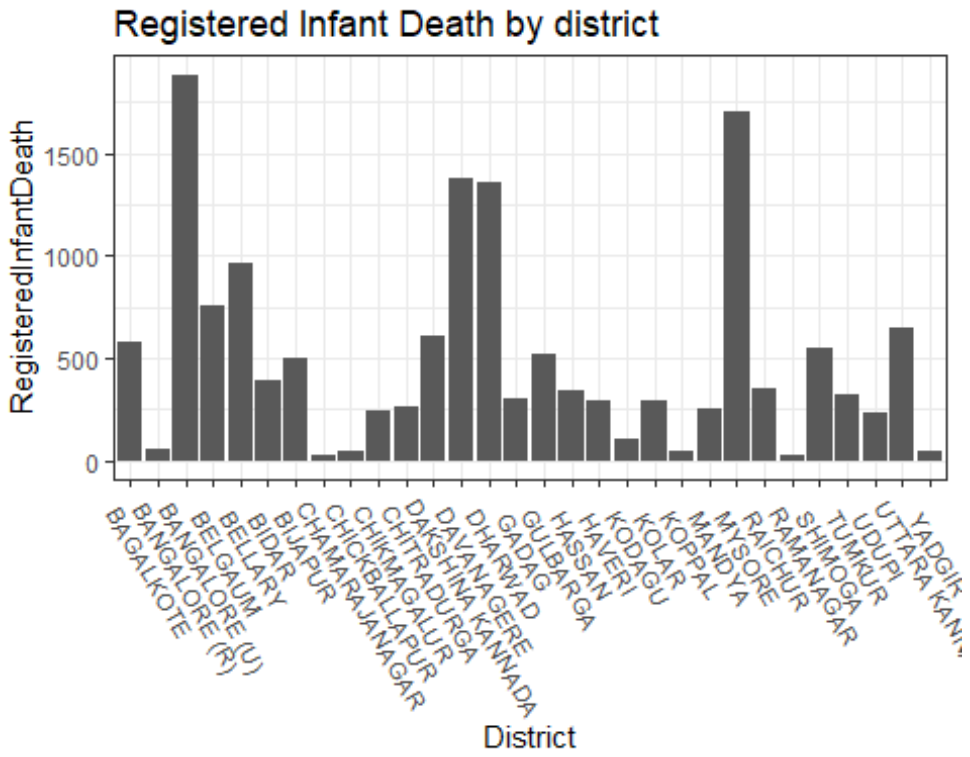
```r
data15 <- read.csv("D:\\Vishal\\III year\\Data Analytics\\Assignment I\\data15.csv", nrows = 30)
data15 <- Filter(function(x)!all(is.na(x)), data15)
tail(data15)
```

```
##          District RegisteredBirth BirthRate RegisteredDeath DeathRate
## 25      RAMANAGAR            8154     13.96            1719      5.50
## 26        SHIMOGA           30719     13.55            6562      5.98
## 27         TUMKUR           31416     14.46            6011      5.92
## 28          UDUPI           14907     13.33            3823      3.96
## 29 UTTARA KANNADA           24387     20.41            3281      4.79
## 30         YADGIR            8417     22.18            1750      5.13
##    RegisteredInfantDeath RegisteredStillBirth StillBirthRate
## 25                    28                   11           0.24
## 26                   549                  269           4.55
## 27                   327                   44           0.75
## 28                   234                  108           2.73
## 29                   652                  127           4.52
## 30                    52                    0           0.00
```

```r
col <- data15 %>%
  select(RegisteredBirth, BirthRate, RegisteredDeath, DeathRate, RegisteredInfantDeath,
         RegisteredStillBirth, StillBirthRate)

d <- colnames(col)

ggplot(data15, aes(x = District)) +
  geom_bar(aes(y = RegisteredBirth), stat="identity", position ="dodge") +
  theme_bw()+
  theme(axis.text.x = element_text(angle=-60, hjust=.1)) +
  labs(title = "Registered Births by district")
```



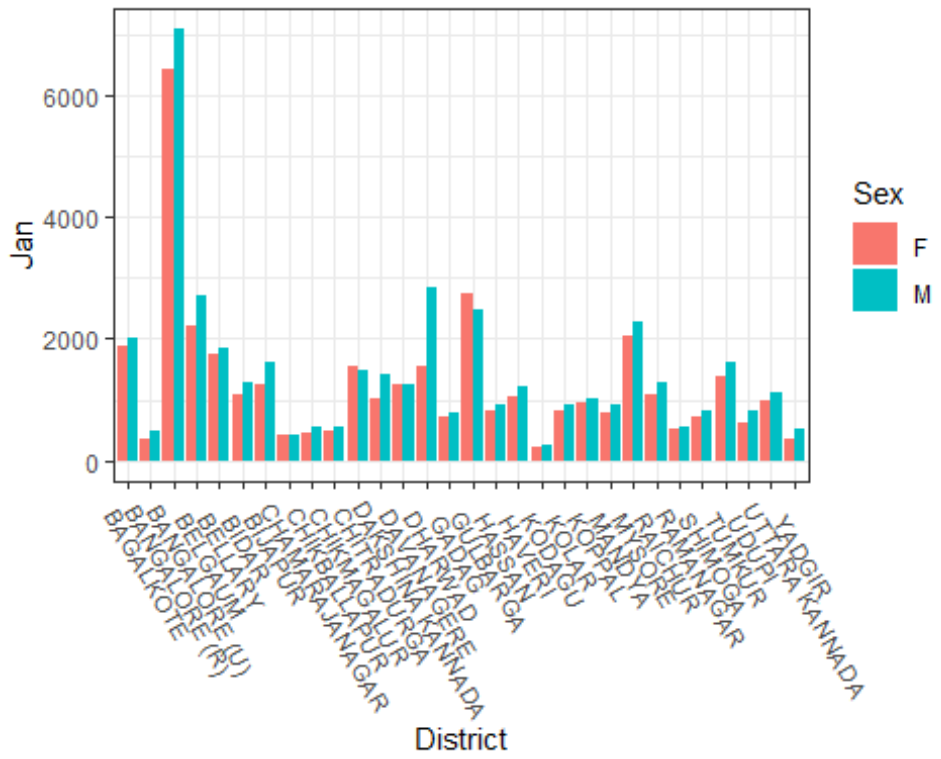

```
cor(data15$BirthRate, data15$DeathRate)

## [1] -0.00997159

ggplot(data15, aes(x = District)) +
  geom_bar(aes(y = RegisteredInfantDeath), stat="identity", position ="dodge"
) +
  theme_bw()+
  theme(axis.text.x = element_text(angle=-60, hjust=.1)) +
  labs(title = "Registered Infant Death by district")
```



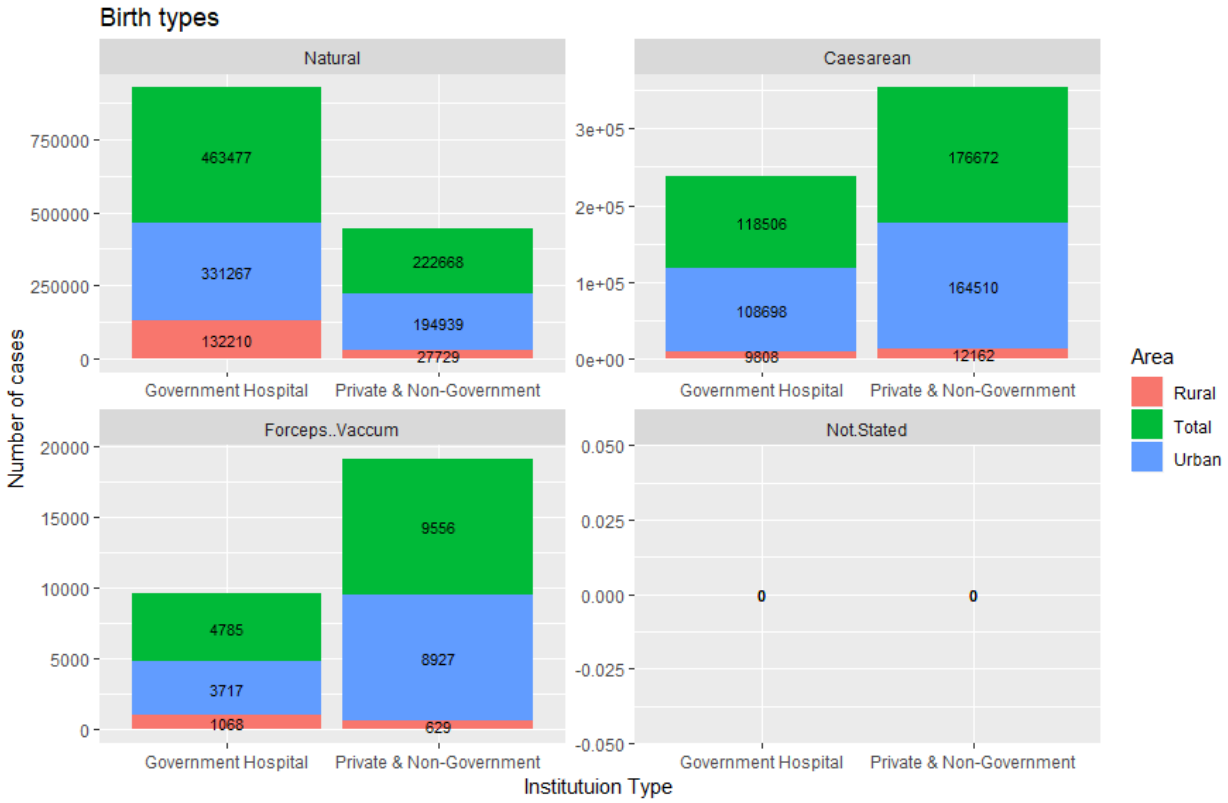

```
data15_2 <- read.csv("D:\\Vishal\\III year\\Data Analytics\\Assignment I\\dat
a15.csv", skip = 36, nrows = 60)

head(data15_2)
```

```
##        District Sex  Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov
## 1     BAGALKOTE   M 2003 1911 2101 1986 2110 1916 2407 2249 2124 2097 2304
## 2     BAGALKOTE   F 1892 1789 1877 1749 1934 1716 2050 1927 1832 2103 1974
## 3 BANGALORE (R)   M  475  481  384  582  369  336  432  406  308  395  302
## 4 BANGALORE (R)   F  351  383  440  539  319  344  303  338  273  422  266
## 5 BANGALORE (U)   M 7070 6345 7520 6531 7029 6744 6928 6858 6996 6491 6778
## 6 BANGALORE (U)   F 6429 5894 7011 6080 6705 6463 6458 6128 6534 6123 6413
##    Dec Total
## 1 2532 25740
## 2 2268 23111
## 3  274  4744
## 4  243  4221
## 5 7090 82380
## 6 6654 76892
```

```
p1 <- ggplot(data15_2, aes(x = District, y = Jan)) +
  geom_bar(aes(fill = Sex), stat="identity", position ="dodge") +
  theme_bw()+
  theme(axis.text.x = element_text(angle=-60, hjust = .1))
p1
```




```
data15_3 <- read.csv("D:\\Vishal\\III year\\Data Analytics\\Assignment I\\dat
a15.csv", skip = 108, nrows = 6, header = TRUE)

data15_3 <- Filter(function(x)! all(is.na(x)), data15_3)

melted <- melt(data15_3, Institution = c("Institution", "Area"))

## Using Institution, Area as id variables

ggplot(melted, aes(x = Institution, y = value, group = 1, label = value)) +
  geom_bar(aes(fill = Area), stat = "identity") +
  facet_wrap(~variable, scales = "free") +
  geom_text(size = 3, position = position_stack(vjust = 0.5)) +
  ylab("Number of cases") + xlab("Institutuion Type") +
  labs(title = "Birth types")
```



data2016.r

Vishal

Sun Feb 17 11:11:30 2019

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

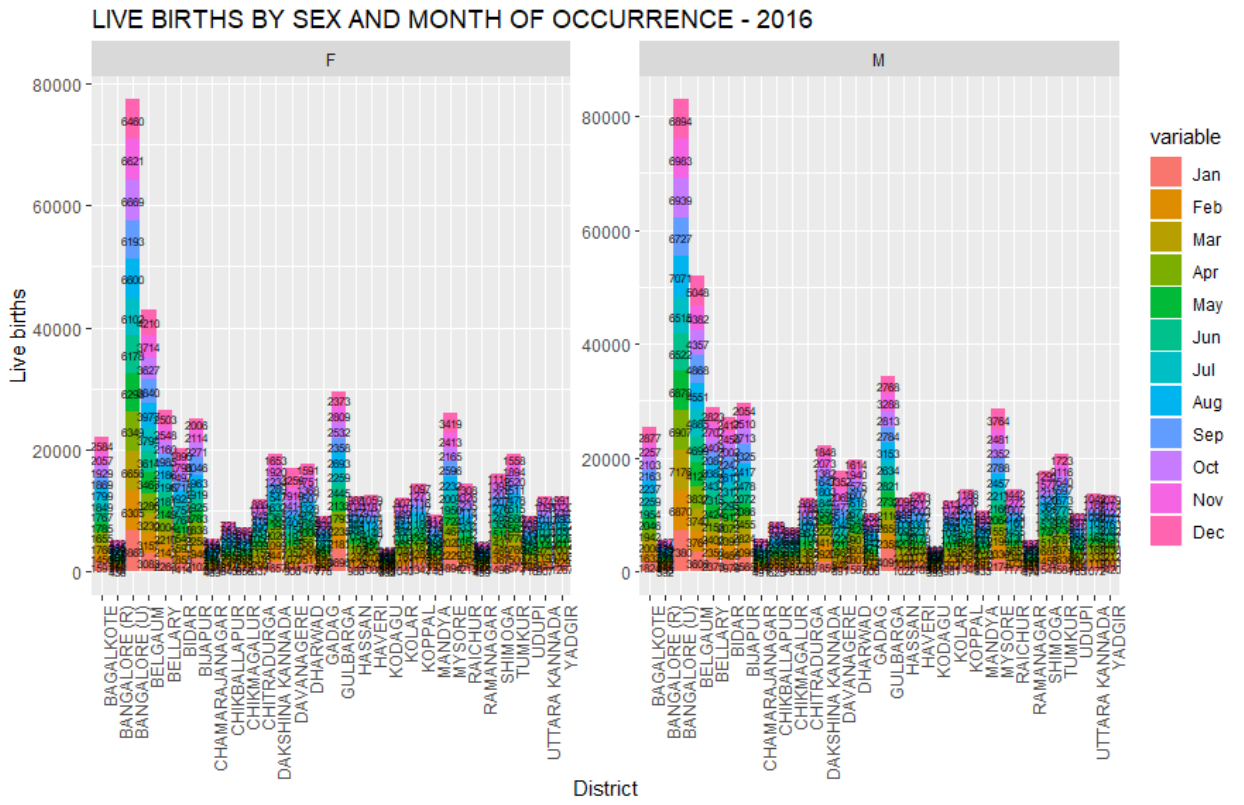
```r
library(ggplot2)
library(reshape2)

data16 <- read.csv("D:\\Vishal\\III year\\Data Analytics\\Assignment I\\data16.csv", nrows = 60)

tail(data16)
```

```
##          District Sex  Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct
## 55          UDUPI   M  765  858  990  867  976  762  772  855  807  847
## 56          UDUPI   F  716  788  843  773  893  678  649  801  713  756
## 57 UTTARA KANNADA   M 1072 1119 1333 1168 1235 1164 1132 1105 1090 1008
## 58 UTTARA KANNADA   F  907  953 1198  973 1149 1061 1034  976 1010  971
## 59         YADGIR   M 1420 1003 1212 1038 1081 1137 1082 1163 1071 1022
## 60         YADGIR   F 1267  929 1106 1005 1012 1028 1050 1052  896  934
##     Nov  Dec  X X.1
## 55  887  803 NA  NA
## 56  808  749 NA  NA
## 57 1164 1042 NA  NA
## 58 1098  974 NA  NA
## 59 1174 1079 NA  NA
## 60 1120  991 NA  NA
```

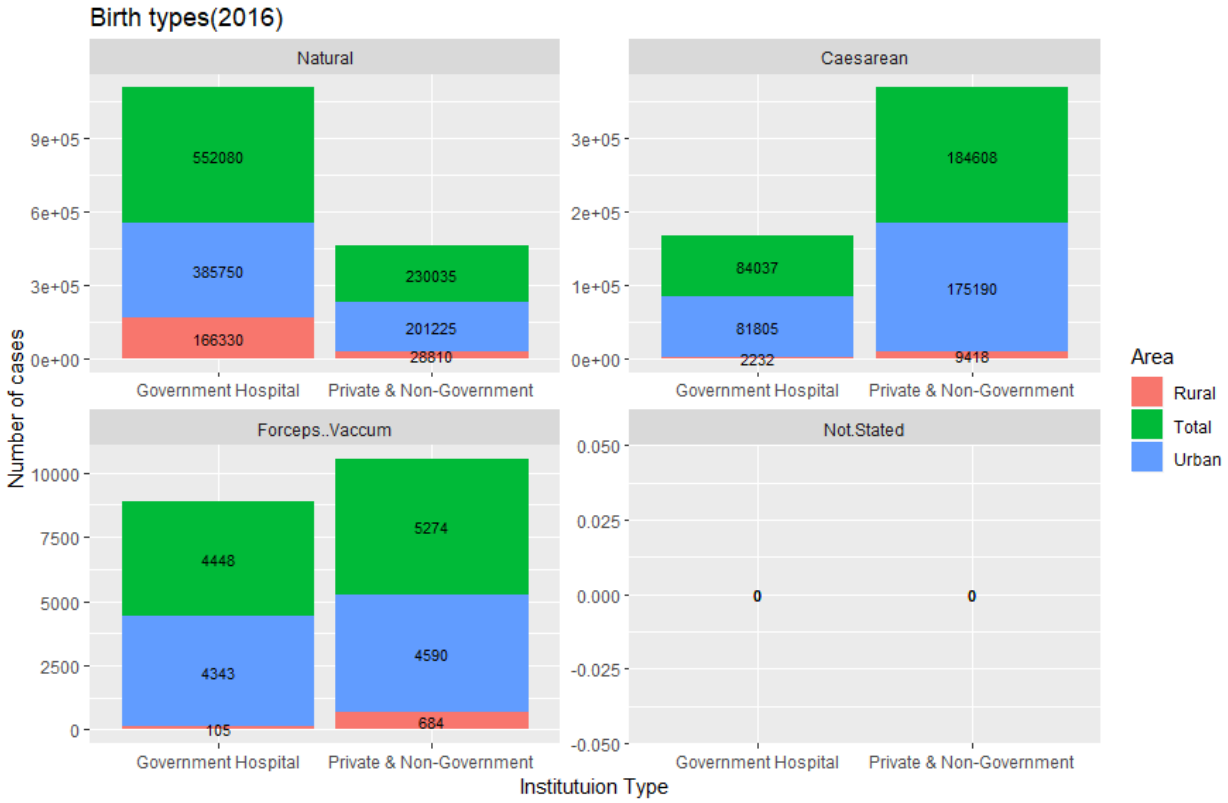
```r
melted <- melt(data16, Months = c("Jan", "Feb", "Mar", "Apr",   "May", "Jun", "Jul",
                                  "Aug", "Sep",   "Oct", "Nov", "Dec"))
```

```
## Using District, Sex, X, X.1 as id variables
```

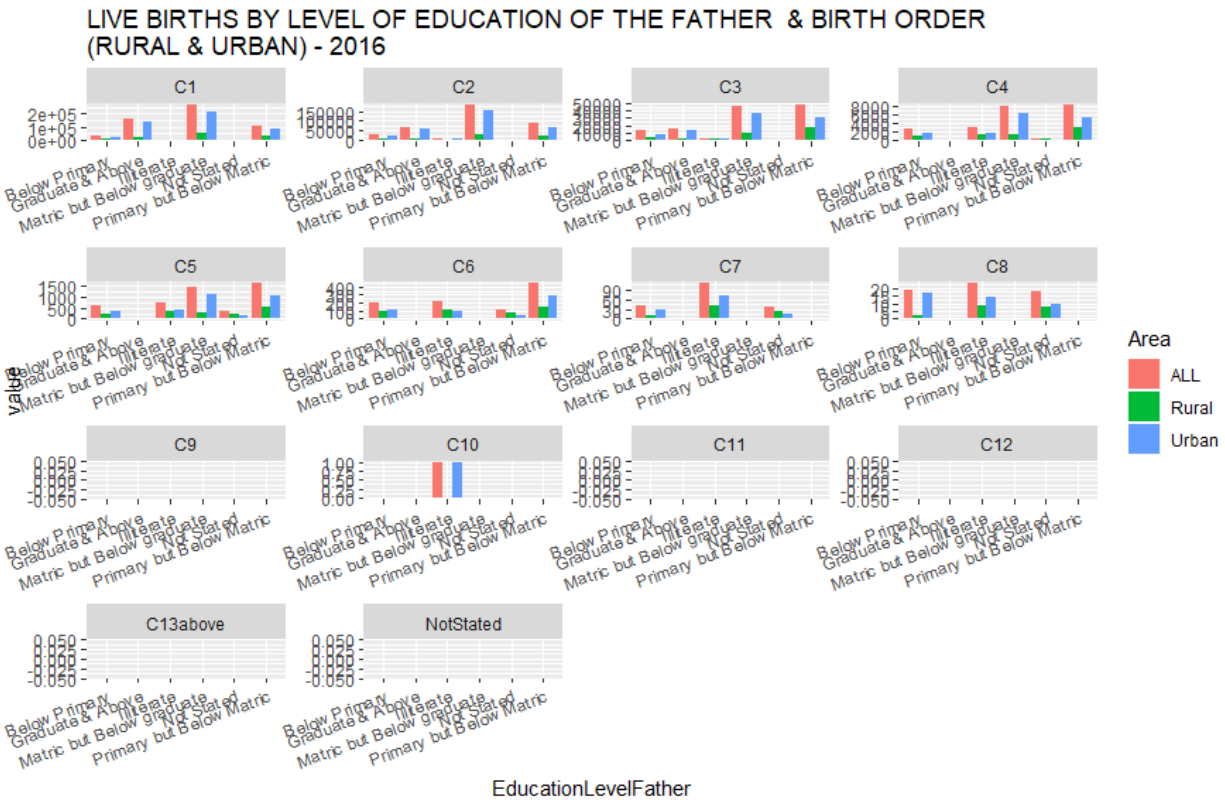
```r
ggplot(melted, aes(x = District, y = value, group = 1, label = value)) +
  geom_bar(aes(fill = variable), stat = "identity") +
  facet_wrap(~Sex, scales = "free") +
  geom_text(size = 2, position = position_stack(vjust = 0.5)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(title = "LIVE BIRTHS BY SEX AND MONTH OF OCCURRENCE - 2016") +
  ylab("Live births")
```




```
data16_2 <- read.csv("D:\\Vishal\\III year\\Data Analytics\\Assignment I\\dat
a16.csv", skip = 64, nrows = 6)

data16_2 <- Filter(function(x)! all(is.na(x)), data16_2)

melted <- melt(data16_2, Institution = c("Institution", "Area"))

## Using Institution, Area as id variables

ggplot(melted, aes(x = Institution, y = value, group = 1, label = value)) +
  geom_bar(aes(fill = Area), stat = "identity") +
  facet_wrap(~variable, scales = "free") +
  geom_text(size = 3, position = position_stack(vjust = 0.5)) +
  ylab("Number of cases") + xlab("Institutuion Type") +
  labs(title = "Birth types(2016)")
```




```
data16_3 <- read.csv("D:\\Vishal\\III year\\Data Analytics\\Assignment I\\dat
a16.csv", skip = 73, nrows = 18)
data16_3 <- Filter(function(x)! all(is.na(x)), data16_3)

melted1 <- melt(data16_3, BirtOrder = c("EducationLevelFather", "Area"))

## Using EducationLevelFather, Area as id variables

ggplot(melted1, aes(x = EducationLevelFather, y = value, label = value)) +
  geom_bar(aes(fill = Area), stat = "identity", position = "dodge") +
  facet_wrap(~variable, scale = "free") +
  theme(axis.text.x = element_text(angle = 20, hjust = 1)) +
  labs(title = "LIVE BIRTHS BY LEVEL OF EDUCATION OF THE FATHER  & BIRTH ORDE
R
(RURAL & URBAN) - 2016")
```



AllYears.r

Vishal

Sun Feb 17 11:14:00 2019

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(ggplot2)
library(reshape2)
library(ggpubr)
```

```
## Loading required package: magrittr
```

```r
dataAll <- read.csv("D:\\Vishal\\III year\\Data Analytics\\Assignment I\\AllYears.csv")

tail(dataAll)
```

```
##    MethodofDelivery Below.15 X15.19 X20.24 X25.29 X30.34 X35.39 X40.44
## 43   Forceps/Vacuum        0    367   3810   3249   1235    276     17
## 44       Not Stated        0      0      0      0      0      0      0
## 45          Natural        0   5965 132733  68831   8515   1532    226
## 46        Caesarean        0    345   6087   3765   1166    267     19
## 47   Forceps/Vacuum        0     51    432    239     53     11      3
## 48       Not Stated        0      0      0      0      0      0      0
##    X45...Above Not.Stated  Area Year
## 43           2          0 Urban 2016
## 44           0          0 Urban 2016
## 45          99          0 Rural 2016
## 46           1          0 Rural 2016
## 47           0          0 Rural 2016
## 48           0          0 Rural 2016
```

```r
meltedData <- melt(dataAll)
```

```
## Using MethodofDelivery, Area as id variables
```

```r
filter11 <- dataAll %>%
  filter(Year == 2011)

filter14 <- dataAll %>%
  filter(Year == 2014)

filter15 <- dataAll %>%
  filter(Year == 2015)

filter16 <- dataAll %>%
  filter(Year == 2016)

melt11 <- melt(filter11)
```

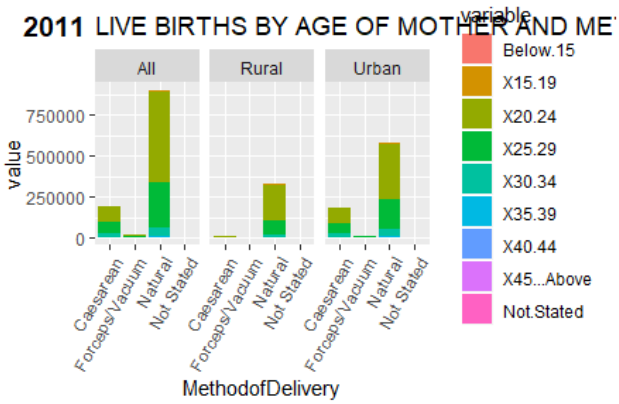
```
## Using MethodofDelivery, Area as id variables
```

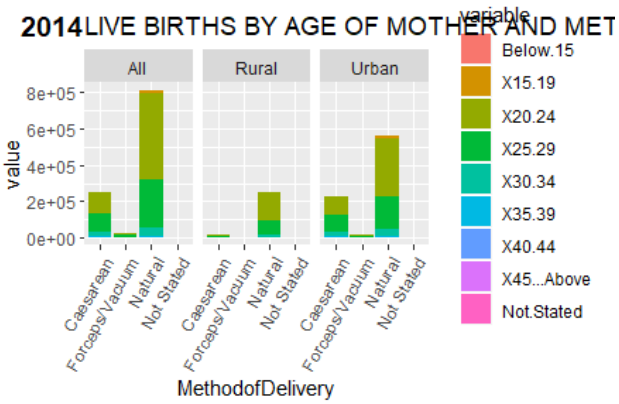
```r
melt14 <- melt(filter14)
```

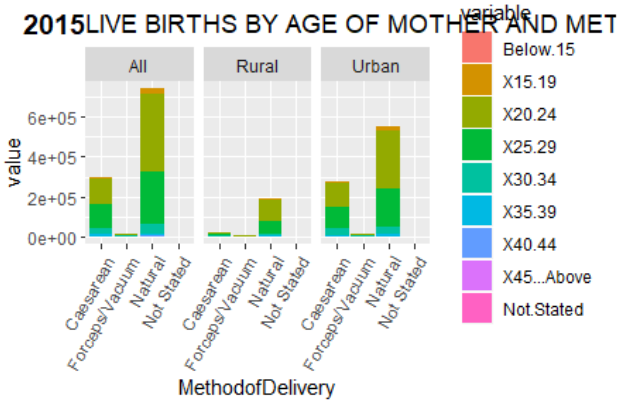
```
## Using MethodofDelivery, Area as id variables
```

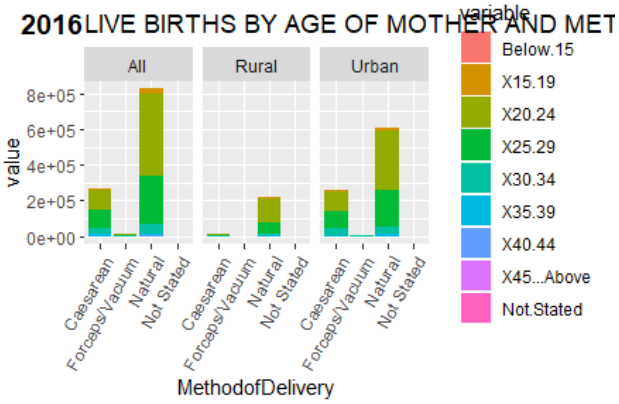
```r
melt15 <- melt(filter15)
```

```r
## Using MethodofDelivery, Area as id variables

melt16 <- melt(filter16)

## Using MethodofDelivery, Area as id variables

melt11_f <- melt11 %>%
  filter(value != 2011)

melt14_f <- melt14 %>%
  filter(value != 2014)

melt15_f <- melt15 %>%
  filter(value != 2015)

melt16_f <- melt16 %>%
  filter(value != 2016)

y1 <- ggplot(melt11_f, aes(x = MethodofDelivery, y = value, label = value)) +
  geom_bar(aes(fill = variable), stat = "identity")+
  facet_wrap(~Area) +
  theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
  labs(title = "LIVE BIRTHS BY AGE OF MOTHER AND METHOD OF DELIVERY  (RURAL &
URBAN) -  2011")

y2 <- ggplot(melt14_f, aes(x = MethodofDelivery, y = value, label = value)) +
  geom_bar(aes(fill = variable), stat = "identity")+
  facet_wrap(~Area) +
  theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
  labs(title = "LIVE BIRTHS BY AGE OF MOTHER AND METHOD OF DELIVERY  (RURAL &
URBAN) -  2014")

y3 <- ggplot(melt15_f, aes(x = MethodofDelivery, y = value, label = value)) +
  geom_bar(aes(fill = variable), stat = "identity")+
  facet_wrap(~Area) +
  theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
  labs(title = "LIVE BIRTHS BY AGE OF MOTHER AND METHOD OF DELIVERY  (RURAL &
URBAN) -  2015")

y4 <- ggplot(melt16_f, aes(x = MethodofDelivery, y = value, label = value)) +
  geom_bar(aes(fill = variable), stat = "identity")+
  facet_wrap(~Area) +
  theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
  labs(title = "LIVE BIRTHS BY AGE OF MOTHER AND METHOD OF DELIVERY  (RURAL &
URBAN) -  2016")


ggarrange(y1, y2, y3, y4,
          labels = c("2011", "2014", "2015", "2016"),
          ncol = 2, nrow = 2)
```


2011 LIVE BIRTHS BY AGE OF MOTHER AND ME
All
Rural
Urban
value
750000
500000
250000
0
Caesarean
Forceps/Vacuum
Natural
Not Stated
MethodofDelivery
variable
Below.15
X15.19
X20.24
X25.29
X30.34
X35.39
X40.44
X45...Above
Not.Stated


2014 LIVE BIRTHS BY AGE OF MOTHER AND MET
All
Rural
Urban
value
8e+05
6e+05
4e+05
2e+05
0e+00
Caesarean
Forceps/Vacuum
Natural
Not Stated
MethodofDelivery
variable
Below.15
X15.19
X20.24
X25.29
X30.34
X35.39
X40.44
X45...Above
Not.Stated


2015 LIVE BIRTHS BY AGE OF MOTHER AND MET
All
Rural
Urban
value
6e+05
4e+05
2e+05
0e+00
Caesarean
Forceps/Vacuum
Natural
Not Stated
MethodofDelivery
variable
Below.15
X15.19
X20.24
X25.29
X30.34
X35.39
X40.44
X45...Above
Not.Stated


2016 LIVE BIRTHS BY AGE OF MOTHER AND MET
All
Rural
Urban
value
8e+05
6e+05
4e+05
2e+05
0e+00
Caesarean
Forceps/Vacuum
Natural
Not Stated
MethodofDelivery
variable
Below.15
X15.19
X20.24
X25.29
X30.34
X35.39
X40.44
X45...Above
Not.Stated