
**CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY(A)**

**Gandipet,Hyderabad-500075**

**2022-2023**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**
**INTRODUCTION TO INFERENCE AND INTERPRETATION (20AMC06)**
**PROJECT REPORT**
**ON**
# ANALYSIS OF INDIAN SUICIDE CASES

**Submitted by:**

Team No: 3

| | | |
|---|---|---|
| 160121729005 | - | Anushka Dodla |
| 160121729015 | - | Varshita Pokala |
| 160121729018 | - | Yarramshetty Charitha Sri |
| 160121729019 | - | Akhil Vanapalli |
| 160121729020 | - | Ali Hasan |

**Under the guidance of:**
Dr. M.Swamy Das
Smt. Ch.Vijaya Lakshmi

# PEER ASSESSMENT

Team No: 3

| S.No | Roll Number | Name of the student | Role | Marks | | | | | | Remarks / Justification |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | a (5) | b (5) | c (10) | d (5) | e (5) | Total (30) | |
| 1 | 160121729019 | Akhil Vanapalli | Lead | 5 | 5 | 9 | 3 | 5 | 27 | Participation as an individual was great, but could've coordinated the team in a better way. |
| 2 | 160121729005 | Anushka Dodla | Member | 5 | 5 | 9 | 4 | 5 | 28 | Very good in the technical department, and has excellent problem solving skills. |
| 3 | 160121729015 | Varshita Pokala | Member | 5 | 5 | 9 | 4 | 5 | 28 | Really supportive as a team member and is very inquisitive. |
| 4 | 160121729018 | Y.Charitha Sri | Member | 5 | 5 | 9 | 4 | 4.5 | 27.5 | Always down to learn something new and is really good at interaction with other team members. |
| 5 | 160121729020 | Ali Hasan | Member | 5 | 5 | 9 | 5 | 5 | 29 | Very keen to work and is very good with deadlines, and is really helpful. |

**Peer assessment guidelines:**

a) Understanding the basic concepts of R and familiarity with RStudio 5M
b) Application of R concepts and visualization tools in real time applications 5M
c) Interpretation of data, analysis and inference 10M
d) Communication (information gathering, slides and report preparation) 5M
e) Participation as an individual and team member 5M

# LIST OF CONTENTS

## OBJECTIVES AND MOTIVATION:

The goal of this project is to analyse and interpret the various aspects of suicide cases in India, in order to educate and spread awareness among our audience about such a serious issue which is plaguing our society.

Our motivation behind selecting this topic was a startling statistic released by the NCRB (The National Crime Records Bureau) which stated that a total of 1,64,033 suicides were reported in the country in 2021, which is the highest ever recorded in the country since the inception of reporting of suicides by the NCRB in 1967. This was extremely depressing and concerning. Hence, through this project, our group has decided to contribute our part and help in alleviating this issue, by spreading awareness and highlighting the different aspects of suicide cases in India.

## DATASET DESCRIPTION:

The dataset in consideration contains yearly suicide details of all the states / Union Terriotories of India by various parameters. The dataset has been published on the Kaggle website and is under the CC (creative commons) license, and hence open to use for the public.

Parameters considered are as follows:

- State
- Year
- Gender
- Age group
- Suicide causes
- Education status
- By means adopted
- Professional profile
- Social status

**LIST OF PACKAGES:**

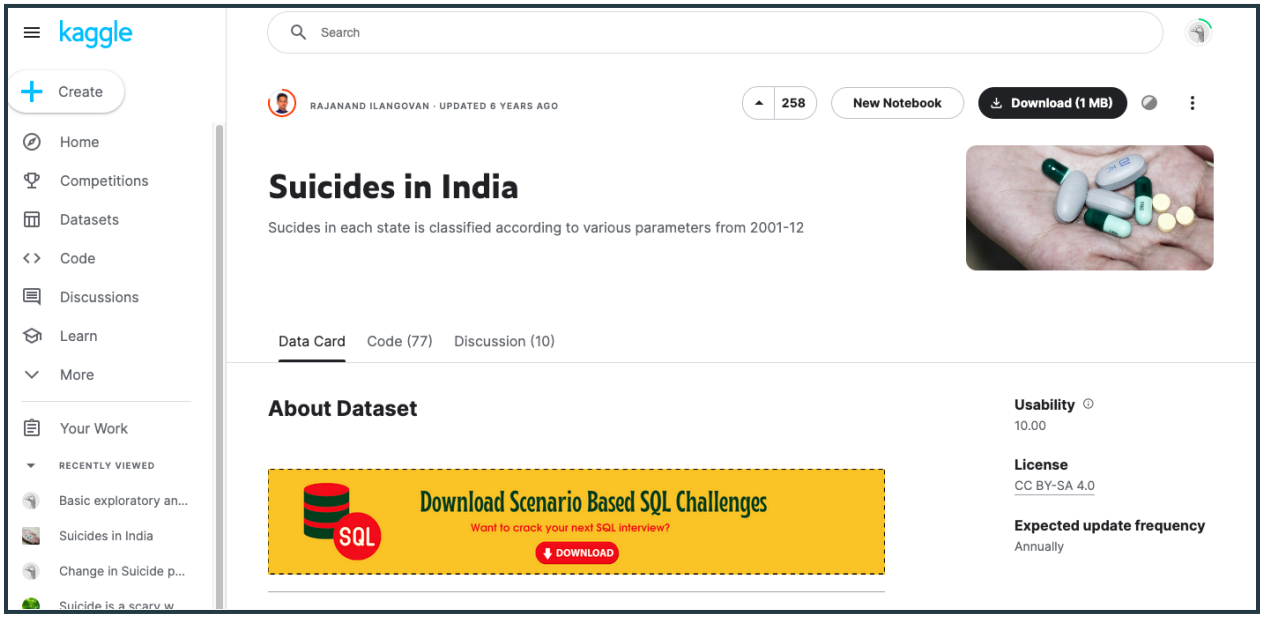
For our experimentation, the following packages were installed and utilized:

1. ggplot2:ggplot2 is a R package dedicated to data visualization. It can greatly improve the quality and aesthetics of your graphics, and will make you much more efficient in creating them. ggplot2 allows you to build almost any type of chart.
2. readr:provides a fast and friendly way to read rectangular data from delimited files, such as comma-separated values (CSV) and tab-separated values (TSV).
3. dplyr: dplyr is a grammar of data manipulation, providing a consistent set of verbs that help you solve the most common data manipulation challenges:
    - mutate() adds new variables that are functions of existing variables
    - select() picks variables based on their names.
    - filter() picks cases based on their values.
    - summarize() reduces multiple values down to a single summary.
    - arrange() changes the ordering of the rows.
    - These all combine naturally with group_by() which allows you to perform any operation “by group”.
4. tidyverse: tidyr is a package that makes it easy to “tidy” your data. Tidy data is data that’s easy to work with: it’s easy to munge (with dplyr), visualize (with ggplot2) and model. The two most important properties of tidy data are:
    - Each column is a variable.
    - Each row is an observation.
    - Arranging your data in this way makes it easier to work with because you have a consistent way of referring to variables (as column names) and observations (as row indices). When using tidy data and tidy tools, you spend less time worrying about how to feed the output from one function into the input of another, and more time answering your questions about the data.
5. ggrepel:ggrepel provides geoms for ggplot2 to repel overlapping text labels.Text labels repel away from each other, away from data points, and away from edges of the plotting area.
    - geom_text_repel()
    - geom_label_repel()

**EXPERIMENTATION:**

First we have downloaded the required dataset from the Kaggle website:

https://www.kaggle.com/datasets/rajanand/suicides-in-india




**PRELIMINARY DATA ANALYSIS** :

Installing the necessary packages and reading the data:

Sample code:

```
# INSTALLING PACKAGES

install.packages('ggplot2')

library(ggplot2)

install.packages('readr')

library(readr)

install.packages('dplyr')

library(dplyr)

install.packages('tidyverse')
```

```
library(tidyverse)

install.packages('ggrepel')

library(ggrepel)

# READING THE DATA

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

suicide_data
```

Executed code:

```
> suicide_data
            State Year Type_code                                      Type Gender Age_group Total
1   A & N Islands 2001    Causes                      Illness (Aids/STD) Female      0-14     0
2   A & N Islands 2001    Causes  Bankruptcy or Sudden change in Economic Female      0-14     0
3   A & N Islands 2001    Causes  Cancellation/Non-Settlement of Marriage Female      0-14     0
4   A & N Islands 2001    Causes        Physical Abuse (Rape/Incest Etc.) Female      0-14     0
5   A & N Islands 2001    Causes                           Dowry Dispute Female      0-14     0
6   A & N Islands 2001    Causes                        Family Problems Female      0-14     0
7   A & N Islands 2001    Causes      Ideological Causes/Hero Worshipping Female      0-14     0
8   A & N Islands 2001    Causes                  Other Prolonged Illness Female      0-14     0
9   A & N Islands 2001    Causes                         Property Dispute Female      0-14     0
10  A & N Islands 2001    Causes                Fall in Social Reputation Female      0-14     0
11  A & N Islands 2001    Causes                   Illegitimate Pregnancy Female      0-14     0
12  A & N Islands 2001    Causes                   Failure in Examination Female      0-14     0
13  A & N Islands 2001    Causes                  Insanity/Mental Illness Female      0-14     0
14  A & N Islands 2001    Causes                             Love Affairs Female      0-14     1
15  A & N Islands 2001    Causes              Professional/Career Problem Female      0-14     0
16  A & N Islands 2001    Causes                                  Divorce Female      0-14     0
17  A & N Islands 2001    Causes                     Drug Abuse/Addiction Female      0-14     0
18  A & N Islands 2001    Causes Not having Children(Barrenness/Impotency Female      0-14     0
19  A & N Islands 2001    Causes                           Causes Not known Female      0-14     0
20  A & N Islands 2001    Causes                               Unemployment Female      0-14     0
21  A & N Islands 2001    Causes            Other Causes (Please Specity) Female      0-14     1
22  A & N Islands 2001    Causes                                  Poverty Female      0-14     0
23  A & N Islands 2001    Causes                    Death of Dear Person Female      0-14     0
24  A & N Islands 2001    Causes                                  Cancer Female      0-14     0
25  A & N Islands 2001    Causes               Suspected/Illicit Relation Female      0-14     0
26  A & N Islands 2001    Causes                                Paralysis Female      0-14     0
27  A & N Islands 2001    Causes                         Property Dispute   Male      0-14     0
28  A & N Islands 2001    Causes                             Unemployment   Male      0-14     0
29  A & N Islands 2001    Causes                                  Poverty   Male      0-14     0
30  A & N Islands 2001    Causes                           Family Problems   Male      0-14     0
```

Printing the dimensions of the dataset:

Sample code:

```
# DIMENSIONS
```

dim(suicide_data)

Executed code:

```
> # DIMENSIONS
> dim(suicide_data)
[1] 237519      7
>
```

Printing the summary of the data:

Sample code:

# SUMMARY

summary(suicide_data)

Executed code:

```
> # SUMMARY
> summary(suicide_data)
    State                Year        Type_code             Type              Gender
 Length:237519      Min.   :2001   Length:237519      Length:237519      Length:237519
 Class :character   1st Qu.:2004   Class :character   Class :character   Class :character
 Mode  :character   Median :2007   Mode  :character   Mode  :character   Mode  :character
                    Mean   :2007
                    3rd Qu.:2010
                    Max.   :2012
  Age_group            Total
 Length:237519      Min.   :    0.00
 Class :character   1st Qu.:    0.00
 Mode  :character   Median :    0.00
                    Mean   :   55.03
                    3rd Qu.:    6.00
                    Max.   :63343.00
```

Printing the first and last 50 observations:

Sample code:

# PRINTING FIRST 50 OBSERVATIONS

head(suicide_data,n=50)

# PRINTING LAST 50 OBSERVATIONS

tail(suicide_data,n=50)

Executed code:

```
> # PRINTING FIRST 50 OBSERVATIONS
> head(suicide_data,n=50)
           State Year Type_code                                    Type Gender Age_group Total
1  A & N Islands 2001    Causes                    Illness (Aids/STD) Female      0-14     0
2  A & N Islands 2001    Causes  Bankruptcy or Sudden change in Economic Female      0-14     0
3  A & N Islands 2001    Causes  Cancellation/Non-Settlement of Marriage Female      0-14     0
4  A & N Islands 2001    Causes        Physical Abuse (Rape/Incest Etc.) Female      0-14     0
5  A & N Islands 2001    Causes                          Dowry Dispute Female      0-14     0
6  A & N Islands 2001    Causes                       Family Problems Female      0-14     0
7  A & N Islands 2001    Causes      Ideological Causes/Hero Worshipping Female      0-14     0
8  A & N Islands 2001    Causes                 Other Prolonged Illness Female      0-14     0
9  A & N Islands 2001    Causes                       Property Dispute Female      0-14     0
10 A & N Islands 2001    Causes               Fall in Social Reputation Female      0-14     0
11 A & N Islands 2001    Causes                  Illegitimate Pregnancy Female      0-14     0
12 A & N Islands 2001    Causes                  Failure in Examination Female      0-14     0
13 A & N Islands 2001    Causes                 Insanity/Mental Illness Female      0-14     0
14 A & N Islands 2001    Causes                             Love Affairs Female      0-14     1
15 A & N Islands 2001    Causes             Professional/Career Problem Female      0-14     0
16 A & N Islands 2001    Causes                                 Divorce Female      0-14     0
17 A & N Islands 2001    Causes                   Drug Abuse/Addiction Female      0-14     0
18 A & N Islands 2001    Causes Not having Children(Barrenness/Impotency Female      0-14     0
19 A & N Islands 2001    Causes                         Causes Not known Female      0-14     0
20 A & N Islands 2001    Causes                             Unemployment Female      0-14     0
```

```
49 A & N Islands 2001    Causes  Bankruptcy or Sudden change in Economic   Male      0-14     0
50 A & N Islands 2001    Causes                 Insanity/Mental Illness   Male      0-14     0
> # PRINTING LAST 50 OBSERVATIONS
> tail(suicide_data,n=50)
             State Year            Type_code                                    Type Gender Age_group
237470 West Bengal 2012 Professional_Profile                                 Student Female     45-59
237471 West Bengal 2012 Professional_Profile                              Unemployed Female     45-59
237472 West Bengal 2012 Professional_Profile                  Service (Government) Female     45-59
237473 West Bengal 2012 Professional_Profile                     Service (Private) Female     45-59
237474 West Bengal 2012 Professional_Profile                 Professional Activity Female     45-59
237475 West Bengal 2012 Professional_Profile              Public Sector Undertaking Female     45-59
237476 West Bengal 2012 Professional_Profile Self-employed (Business activity) Female     45-59
237477 West Bengal 2012 Professional_Profile Self-employed (Business activity)   Male     45-59
237478 West Bengal 2012 Professional_Profile                              Unemployed   Male     45-59
237479 West Bengal 2012 Professional_Profile                           Retired Person   Male     45-59
237480 West Bengal 2012 Professional_Profile                  Service (Government)   Male     45-59
237481 West Bengal 2012 Professional_Profile                 Professional Activity   Male     45-59
237482 West Bengal 2012 Professional_Profile                Others (Please Specify)   Male     45-59
237483 West Bengal 2012 Professional_Profile              Public Sector Undertaking   Male     45-59
237484 West Bengal 2012 Professional_Profile                              House Wife   Male     45-59
237485 West Bengal 2012 Professional_Profile           Farming/Agriculture Activity   Male     45-59
237486 West Bengal 2012 Professional_Profile                     Service (Private)   Male     45-59
237487 West Bengal 2012 Professional_Profile                                 Student   Male     45-59
237488 West Bengal 2012 Professional_Profile                     Service (Private) Female       60+
```

Performing column operations:

Sample code:

```
# COLUMN NAMES

names(suicide_data)

# ACCESSING SPECIFIC COLUMNS

suicide_data$Age_group

range(suicide_data$Age_group)

suicide_data$Year

range(suicide_data$Year)

suicide_data["Type"]

tail(suicide_data["State"],n=500)

suicide_data[c(500:600),]
```

Executed code:

```
> # COLUMN NAMES
> names(suicide_data)
[1] "State"     "Year"      "Type_code" "Type"      "Gender"    "Age_group" "Total"
>
```

```
> # ACCESSING SPECIFIC COLUMNS
> suicide_data$Age_group
   [1] "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"
  [11] "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"
  [21] "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"
  [31] "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"
  [41] "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"  "0-14"
  [51] "0-14"  "0-14"  "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29"
  [61] "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29"
  [71] "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29"
  [81] "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29"
  [91] "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29" "15-29"
 [101] "15-29" "15-29" "15-29" "15-29" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44"
 [111] "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44"
 [121] "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44"
 [131] "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44"
 [141] "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "30-44"
 [151] "30-44" "30-44" "30-44" "30-44" "30-44" "30-44" "45-59" "45-59" "45-59" "45-59"
 [161] "45-59" "45-59" "45-59" "45-59" "45-59" "45-59" "45-59" "45-59" "45-59" "45-59"
 [171] "45-59" "45-59" "45-59" "45-59" "45-59" "45-59" "45-59" "45-59" "45-59" "45-59"
 [181] "45-59" "45-59" "45-59" "45-59" "45-59" "45-59" "45-59" "45-59" "45-59" "45-59"
```

```
> range(suicide_data$Age_group)
[1] "0-100+" "60+"
>
```

```
> suicide_data$Year
   [1] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
  [19] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
  [37] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
  [55] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
  [73] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
  [91] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
 [109] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
 [127] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
 [145] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
 [163] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
 [181] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
 [199] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
 [217] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
 [235] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
 [253] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
 [271] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
 [289] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
 [307] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
 [325] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
 [343] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
 [361] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
 [379] 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001
```

```
> range(suicide_data$Year)
[1] 2001 2012
>
```

```
2            Bankruptcy or Sudden change in Economic
3            Cancellation/Non-Settlement of Marriage
4                  Physical Abuse (Rape/Incest Etc.)
5                                       Dowry Dispute
6                                    Family Problems
7                  Ideological Causes/Hero Worshipping
8                               Other Prolonged Illness
9                                    Property Dispute
10                         Fall in Social Reputation
11                            Illegitimate Pregnancy
12                            Failure in Examination
13                           Insanity/Mental Illness
14                                       Love Affairs
15                      Professional/Career Problem
16                                           Divorce
17                               Drug Abuse/Addiction
18          Not having Children(Barrenness/Impotency
19                                   Causes Not known
20                                       Unemployment
21                     Other Causes (Please Specity)
22                                           Poverty
23                               Death of Dear Person
24                                           Cancer
```

Creating a contingency table:

Sample code:

suicide.gender.tab<-table(suicide=suicide_data$Gender)

suicide.gender.tab

addmargins(suicide.gender.tab)

Executed code:

```
> # CONTINGENCY TABLE
> suicide.gender.tab<-table(suicide=suicide_data$Gender)
> suicide.gender.tab
suicide
Female   Male
118640 118879
> addmargins(suicide.gender.tab)
suicide
Female   Male    Sum
118640 118879 237519
> mean(suicide_data$Total)
[1] 55.03448
>
```

Selecting 10 random rows from the dataset:

Sample code:

sample_n(suicide_data, 10)

Executed code:

```
> sample_n(suicide_data, 10)
               State Year            Type_code                    Type Gender Age_group Total
1       D & N Haveli 2006              Causes        Property Dispute   Male       60+     0
2          Meghalaya 2004       Means_adopted By Fire/Self Immolation Female     45-59     0
3        Maharashtra 2006       Means_adopted      By Over Alcoholism   Male     45-59    47
4       Chhattisgarh 2002              Causes                 Divorce   Male      0-14     0
5         Puducherry 2011 Professional_Profile             House Wife Female     15-29    45
6        Daman & Diu 2011              Causes            Unemployment   Male      0-14     0
7    Jammu & Kashmir 2002              Causes                  Cancer   Male      0-14     0
8  Arunachal Pradesh 2007 Professional_Profile Others (Please Specify)   Male     15-29     4
9                Goa 2011              Causes        Property Dispute Female     30-44     0
10           Odisha 2007              Causes         Family Problems   Male     30-44   267
>
```

**DATA VISUALIZATION:**

Visualizing how the number of suicide cases changed from 2001 to 2012:

Sample code:

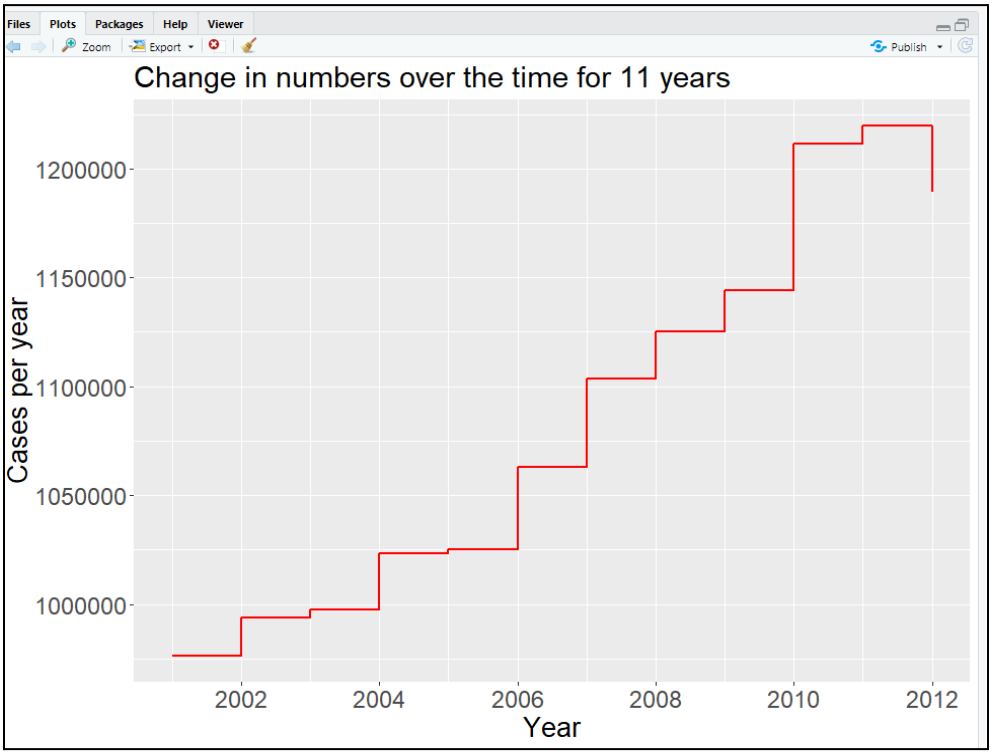
# Plot 1 - How did the numbers of suicide cases change over 2001-2012

cases_over_11_years = suicide_data %>%

select(Year, Total) %>%

```
  arrange(Year) %>%
  group_by(Year) %>%
  summarize(Total =  sum(Total))
options(repr.plot.width = 12, repr.plot.height = 10)# plot dimensions
plot1 = ggplot(data = cases_over_11_years)
plot1 + geom_step(aes(x = Year, y = Total), stat = "identity", size = 1, color = "red") +
  labs( title = "Change in numbers over the time for 11 years",
        x = "Year",y = "Cases per year")+
  theme(axis.text = element_text(size = 18)) +
  theme(axis.title = element_text(size = 20)) +
  theme(plot.title = element_text(size=22)) +
  scale_x_continuous(breaks = ~ axisTicks(., log = FALSE))
```

Executed code:

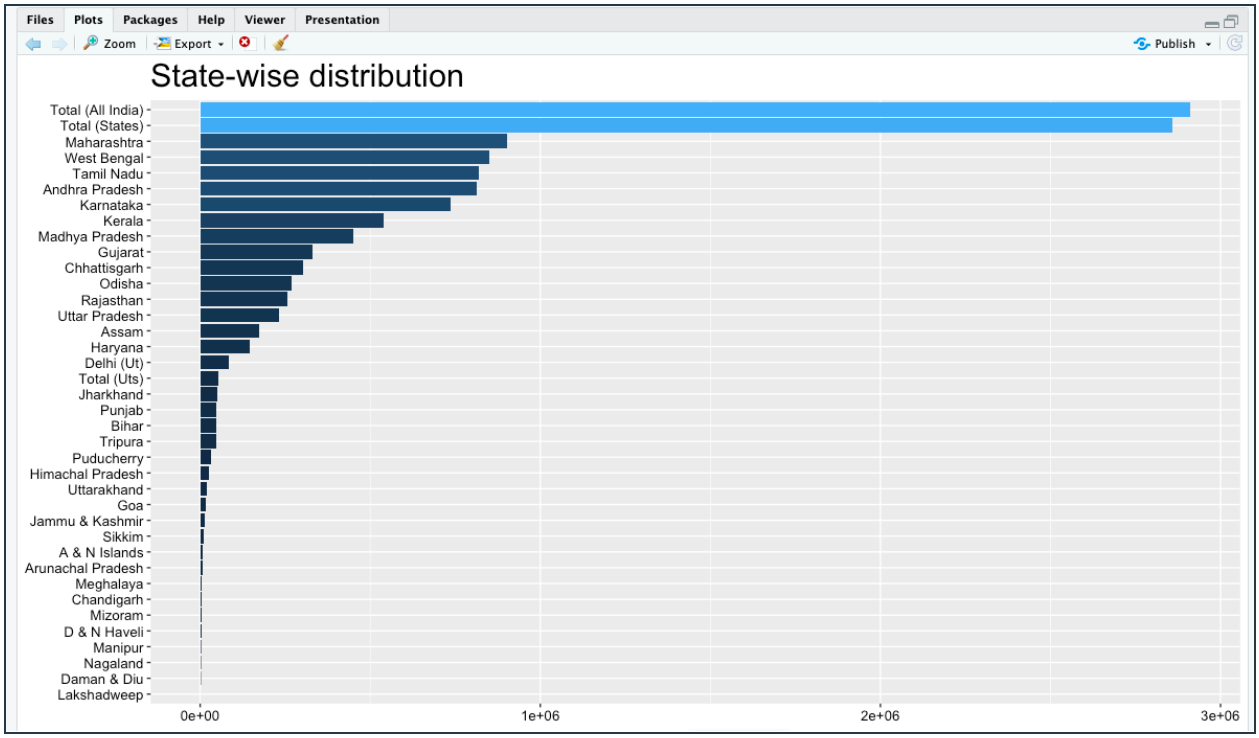


Visualizing state-wise suicide cases:

Sample code:

```
State_wise_df = suicide_data %>%
  select(State, Total) %>%
  group_by(State) %>%
  summarize(Total = sum(Total)) %>%
  arrange(State, -Total)
State_wise_df = data.frame(State_wise_df)
write.csv(State_wise_df,"State Wise.csv")
plot1 = ggplot(data = State_wise_df)
plot1 + geom_col(mapping = aes(x = Total, y = reorder(State, Total), fill = Total)) +
  theme(axis.text = element_text(size = 10, colour = "black")) +
  theme(legend.position = "none") +
  scale_fill_gradient(guide = "colourbar") +
  theme(axis.title = element_blank()) +
  labs(title = "State-wise distribution") +
  theme(plot.title = element_text(size = 23))
```

Executed code:



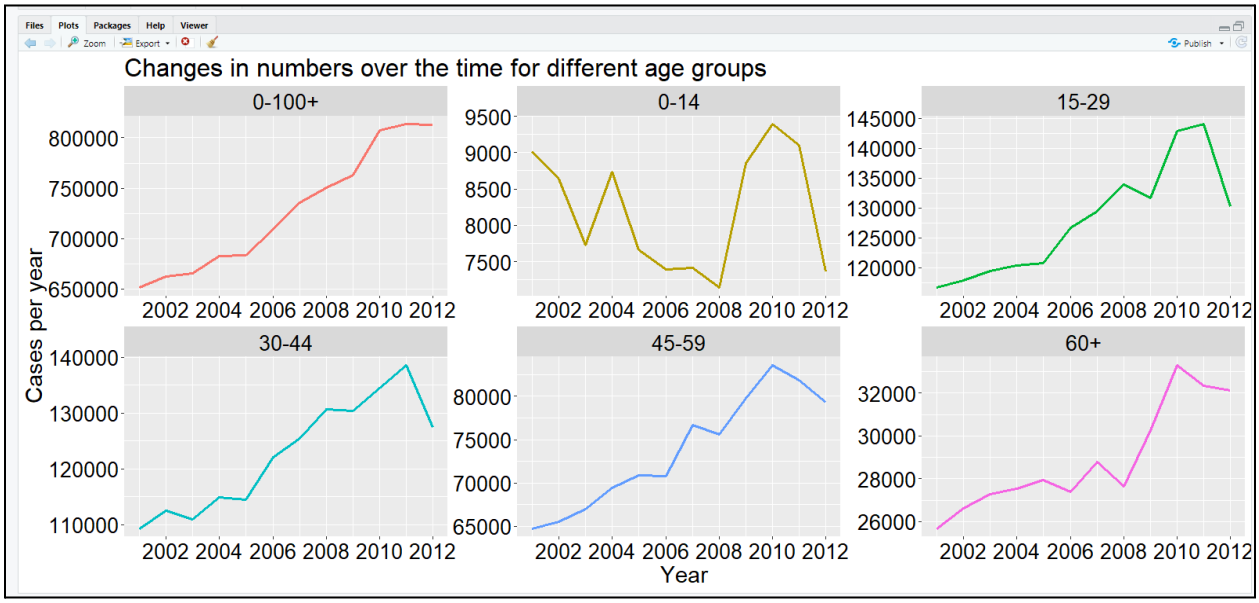

Visualizing how the suicide cases changed for each age group:

Sample code:

```
Cases_by_agegroup = suicide_data %>%
  select(Age_group, Year, Total) %>%
  group_by(Age_group, Year) %>%
  summarize(Total = sum(Total))
options(scipen=999)  # turn off scientific notation like 1e+06
options(repr.plot.width = 20, repr.plot.height = 10) #plot dimensions
plot2 = ggplot(Cases_by_agegroup)
plot2 + geom_line(mapping = aes( x = Year, y = Total, colour = Age_group), stat = "identity",
size = 1.15) +
  facet_wrap(~ Age_group, dir = "h" , scales = "free", nrow = 2, strip.position = "top")+
  theme(strip.text.x = element_text(size = 20, colour = "black"))+
```

```
  theme(legend.position = "none") +
  labs( title = "Changes in numbers over the time for different age groups",
        x = "Year",y = "Cases per year") +
  theme(plot.title = element_text(size=24)) +
  theme(axis.text = element_text(size = 20, colour = "black")) +
  theme(axis.title = element_text(size = 22)) +
  theme(legend.key.size = unit(2, 'cm')) +
  theme(legend.text = element_text(size = 18)) +
  theme(legend.title = element_text(size = 18)) +
  scale_x_continuous(breaks = ~ axisTicks(., log = FALSE))
```

Executed code:



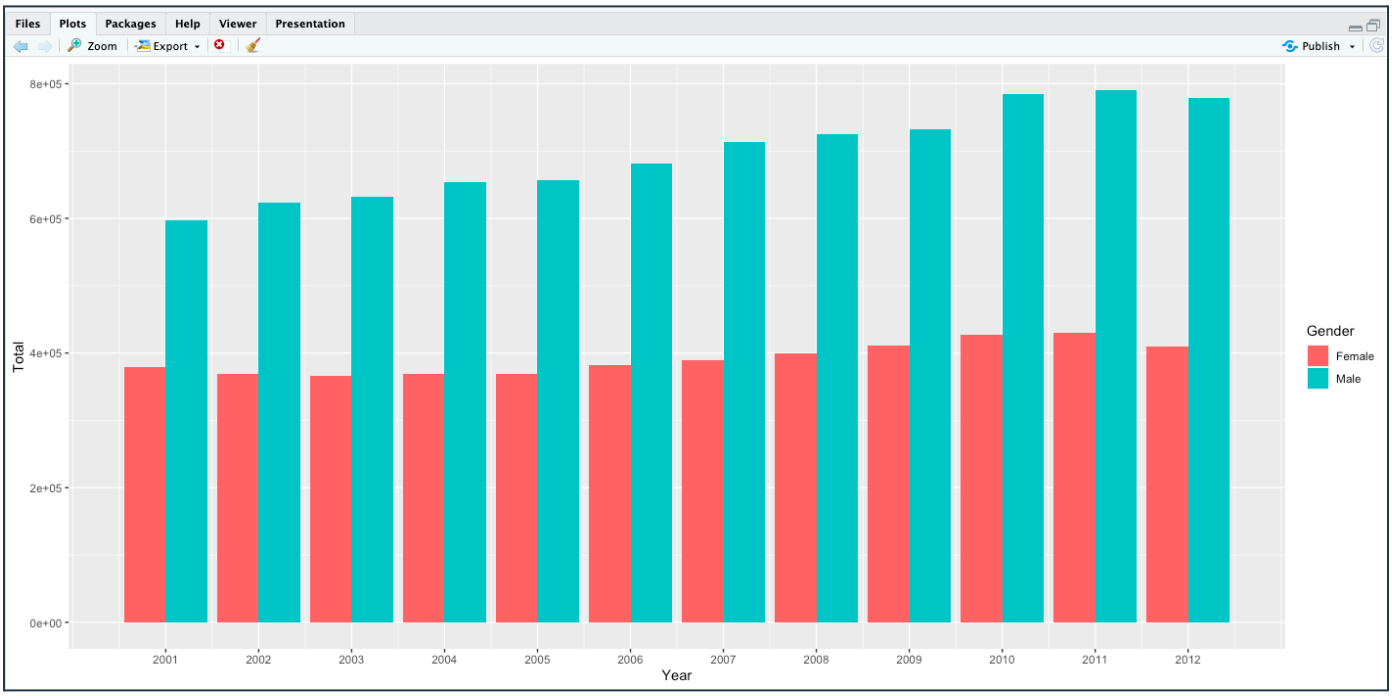

Visualizing yearly changes in suicide cases with gender:

Sample code:

```
Avg_cases_per_gender = suicide_data %>%
  select(Year, Gender, Total) %>%
```

```
  group_by(Year, Gender) %>%
  summarize(Total = sum(Total))
Avg = data.frame(Avg_cases_per_gender)
write.csv(Avg,"Average.csv")
options(repr.plot.width = 10, repr.plot.height = 15)
ggplot(aes(x = Year,y = Total,group = Gender, fill = Gender),data = Avg)+
  geom_bar(position = "dodge", stat = "identity") +
  scale_x_continuous(breaks = seq(2001,2012, by = 1))
```

Executed code:



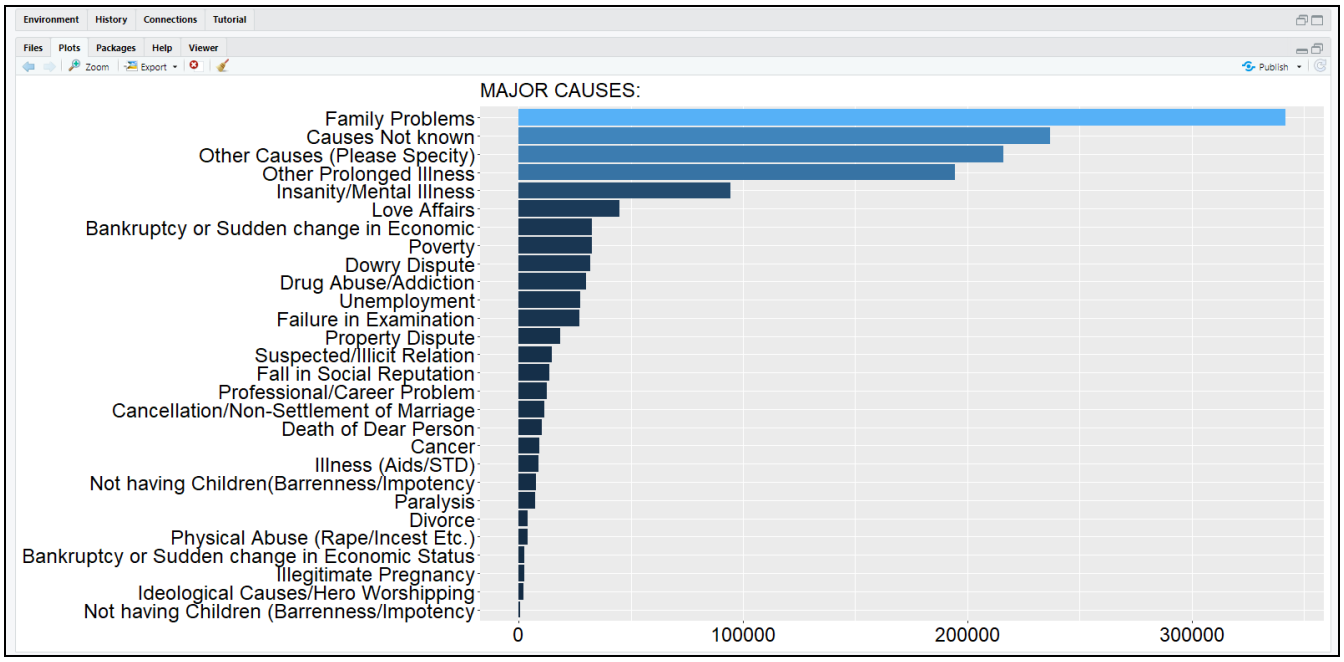

Visualizing the major causes of suicides:

Sample code:

```
Causes_df = Suicides_in_India %>%
            select(Type_code, Type, Total) %>%
            filter(Type_code == "Causes") %>%
```

```
            group_by(Type)%>%
            summarize(Total = sum(Total)) %>%
            arrange (-Total)
Causes_df = data.frame(Causes_df)
options(repr.plot.width = 15, repr.plot.height = 18)
plot4 = ggplot(data = Causes_df)
plot4 + geom_col(mapping = aes(x = Total, y = reorder(Type, Total), fill = Total)) +
      theme(axis.text = element_text(size = 21, colour = "black")) +
      theme(legend.position = "none") +
      scale_fill_gradient(guide = "colourbar") +
      theme(axis.title = element_blank()) +
      labs(title = "MAJOR CAUSES:") +
      theme(plot.title = element_text(size = 23))
```

Executed code:



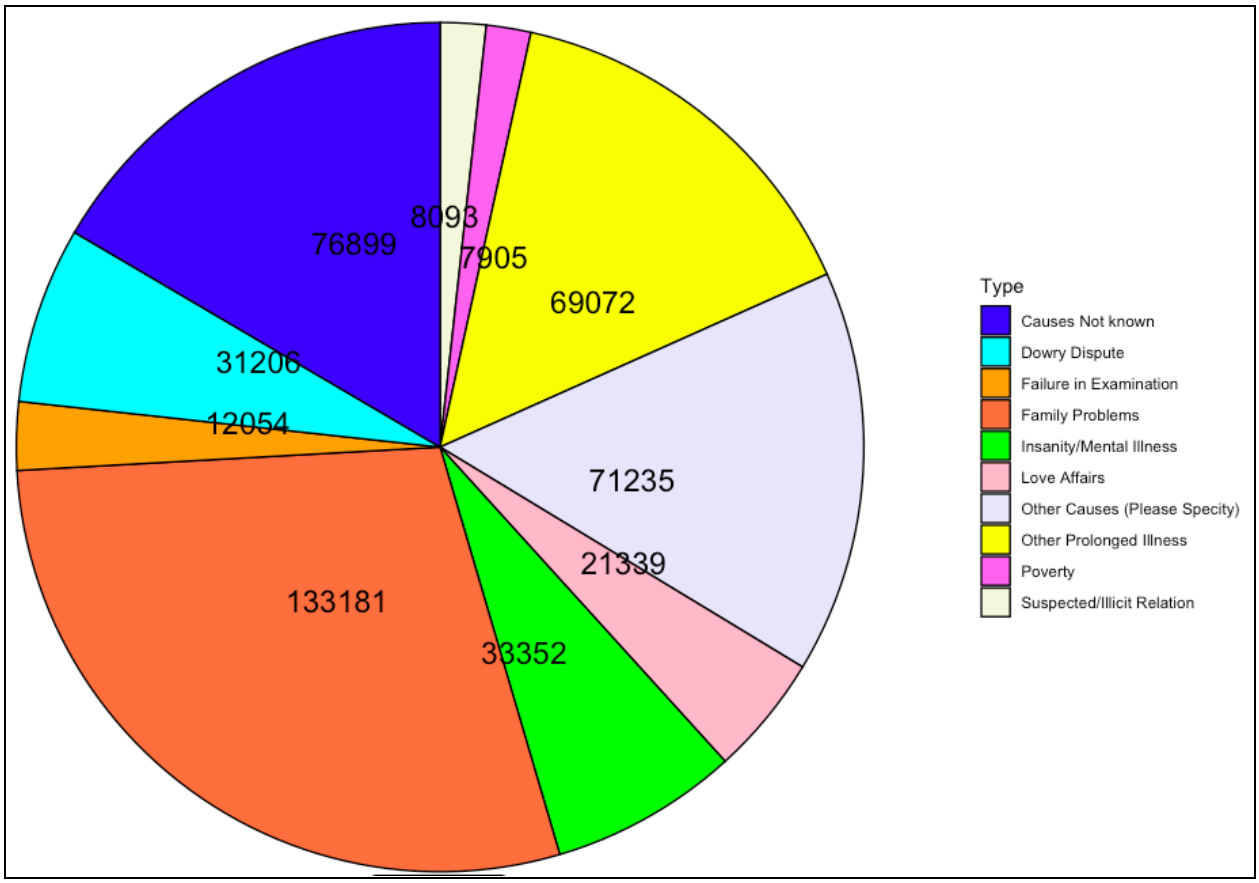

Visualizing the leading causes of suicides in females:

Sample code:

```
Female_cases = suicide_data %>%
  select(Type_code, Type,Gender,Total) %>%
  filter(Type_code == "Causes") %>%
  group_by(Type) %>%
```

```
  subset(Gender=='Female')%>%
  summarize(Total = sum(Total))%>%
  arrange (-Total)
Female_cases = data.frame(Female_cases)
write.csv(Female_cases,"female_cases.csv")
#Leading Causes
leading_causes=c('Family Problems','Causes Not known','Other Causes (Please Specity)',
                 'Other Prolonged Illness','Insanity/Mental Illness','Dowry Dispute',
                 'Love Affairs','Failure in Examination','Suspected/Illicit Relation',
                 'Poverty')
Top10_df = Female_cases %>%
  select(Type, Total) %>%
  filter(Type %in% leading_causes)
Top10_df = data.frame(Top10_df)
write.csv(Top10_df,"Top10.csv")
#Pie chart
Top10_df <- Top10_df %>%
  arrange(desc(Type)) %>%
  mutate(prop = Total / sum(Top10_df$Total) *100) %>%
  mutate(ypos = cumsum(prop)- 0.5*prop )
mycols <- c('blue', 'cyan','orange','coral','green','pink','lavender','yellow',
            'violet','beige')
options(ggrepel.max.overlaps = Inf)
ggplot(Top10_df, aes(x="", y=prop, fill=Type)) +
  geom_bar(stat="identity", width=1, color="black") +
  coord_polar("y", start=0) +
  theme_void() +
  theme(legend.position="none") +
  geom_text_repel(aes(y = ypos, label = Total), color = "black", size=6) +
  scale_fill_manual(values = mycols)+
  theme_void()
```

Executed code:



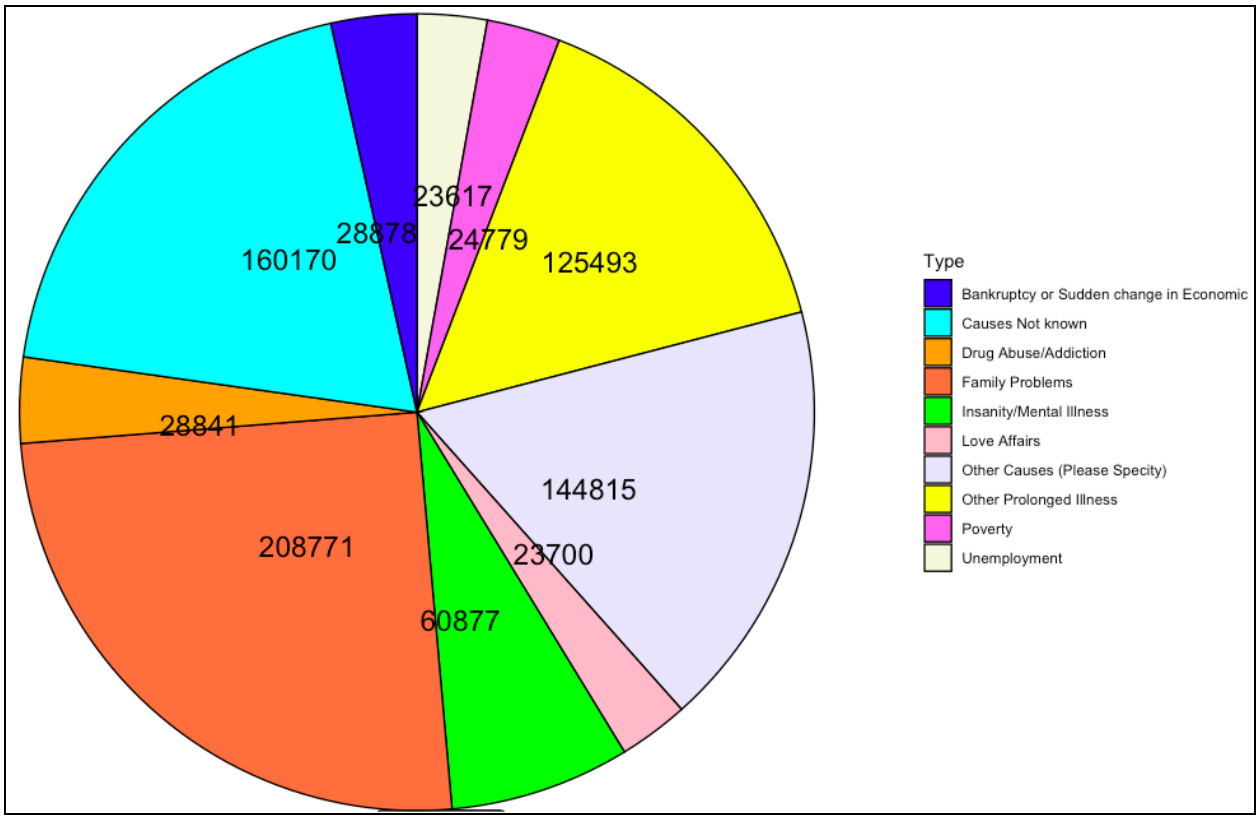

Visualizing leading causes of suicides in males:

Sample code:

```
male_cases = suicide_data %>%
select(Type_code, Type,Gender,Total) %>%
filter(Type_code == "Causes") %>%
group_by(Type) %>%
subset(Gender=='Male')%>%
summarize(Total = sum(Total))%>%
arrange (-Total)
male_cases = data.frame(male_cases)
write.csv(male_cases,"male_cases.csv")
#Leading Causes
leading_causes=c('Family Problems','Causes Not known','Other Causes (Please Specity)',
'Other Prolonged Illness','Insanity/Mental Illness',
'Bankruptcy or Sudden change in Economic','Drug Abuse/Addiction',
'Poverty','Love Affairs','Unemployment')
Top10_df = male_cases %>%
```

```
select(Type, Total) %>%
filter(Type %in% leading_causes)
Top10_df = data.frame(Top10_df)
write.csv(Top10_df,"Top10.csv")
#Pie chart
Top10_df <- Top10_df %>%
  arrange(desc(Type)) %>%
  mutate(prop = Total / sum(Top10_df$Total) *100) %>%
  mutate(ypos = cumsum(prop)- 0.5*prop )
  mycols <- c('blue', 'cyan','orange','coral','green','pink','lavender','yellow',
                          'violet','beige')
  options(ggrepel.max.overlaps = Inf)
  ggplot(Top10_df, aes(x="", y=prop, fill=Type)) +
  geom_bar(stat="identity", width=1, color="black") +
  coord_polar("y", start=0) +
  theme_void() +
  theme(legend.position="none") +
  geom_text_repel(aes(y = ypos, label = Total), color = "black", size=6) +
  scale_fill_manual(values = mycols)+
  theme_void()
```

Executed code:



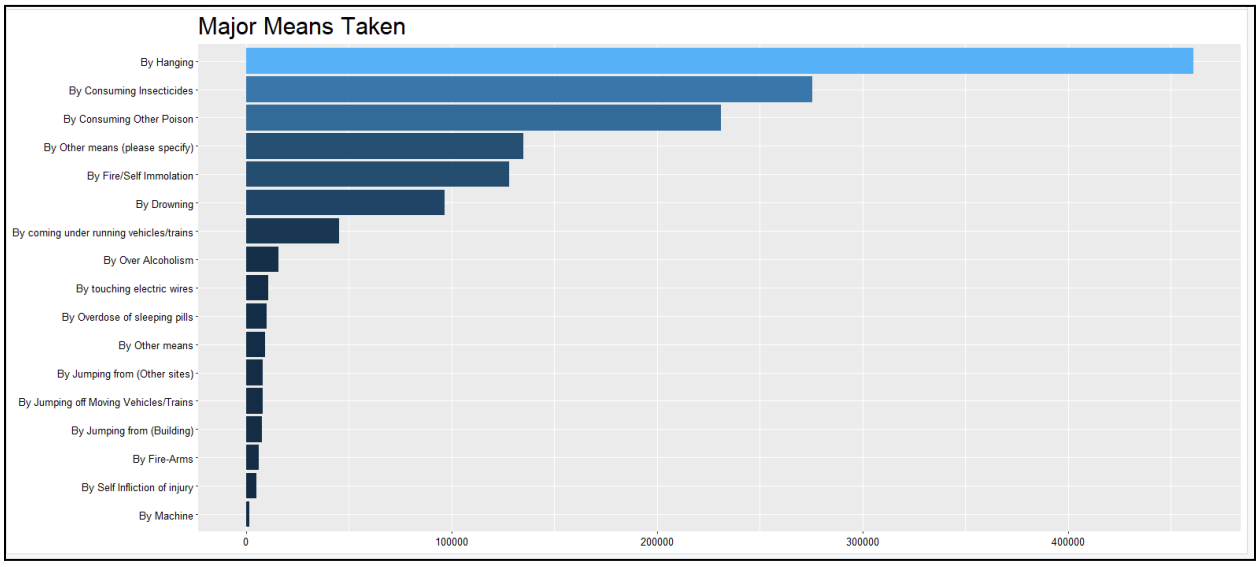

Visualizing the major means by which suicides were committed:

Sample code:

```
Means_df = suicide_data %>%
  select(Type_code, Type, Total) %>%
  filter(Type_code == "Means_adopted") %>%
  group_by(Type)%>%
  summarize(Total = sum(Total)) %>%
  arrange (-Total)
Means_df = data.frame(Means_df)
write.csv(Means_df,"Means.csv")
options(scipen = 999)
options(repr.plot.width = 10, repr.plot.height = 15)
plot1 = ggplot(data = Means_df)
plot1 + geom_col(mapping = aes(x = Total, y = reorder(Type, Total), fill = Total)) +
  theme(axis.text = element_text(size = 10, colour = "black")) +
  theme(legend.position = "none") +
  scale_fill_gradient(guide = "colourbar") +
```

```
  theme(axis.title = element_blank()) +
  labs(title = "Major Means Taken") +
  theme(plot.title = element_text(size = 23))
```

Executed code:



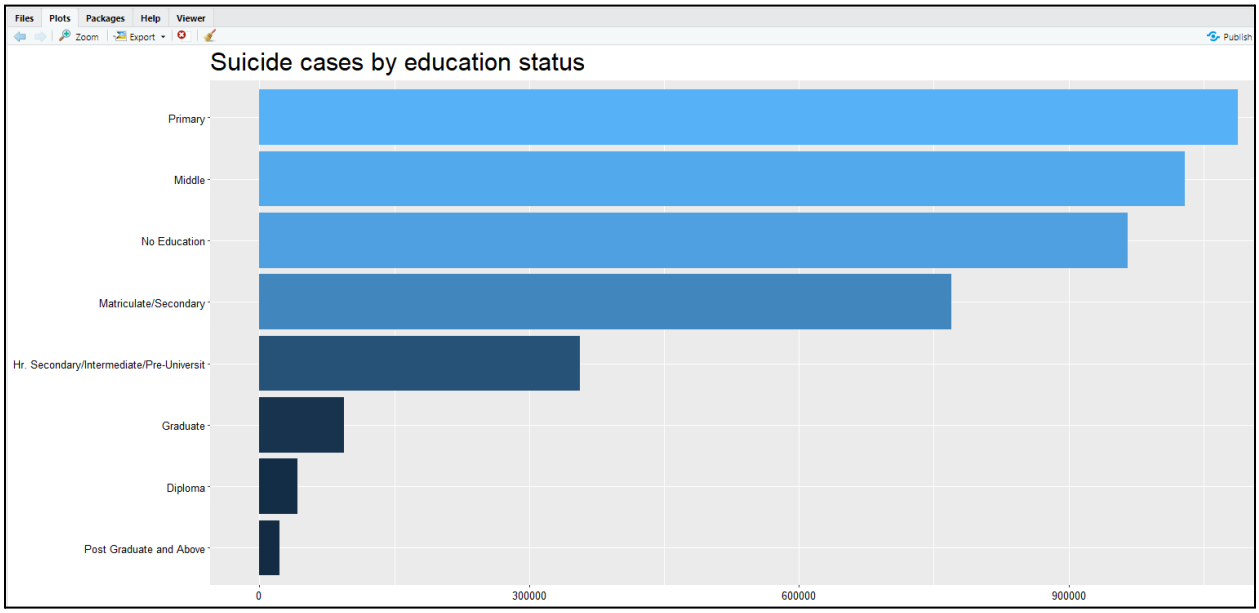

Visualizing suicide cases by education status:

Sample code:

```
education_df = suicide_data %>%
  select(Type_code, Type, Total) %>%
  filter(Type_code == "Education_Status") %>%
  group_by(Type)%>%
  summarize(Total = sum(Total)) %>%
  arrange (-Total)
education_df = data.frame(education_df)
write.csv(education_df,"education.csv")
options(scipen = 999)
options(repr.plot.width = 10, repr.plot.height = 15)
plot1 = ggplot(data = education_df)
plot1 + geom_col(mapping = aes(x = Total, y = reorder(Type, Total), fill = Total)) +
  theme(axis.text = element_text(size = 10, colour = "black")) +
  theme(legend.position = "none") +
  scale_fill_gradient(guide = "colourbar") +
  theme(axis.title = element_blank()) +
```

```
  labs(title = "Suicide cases by education status") +
  theme(plot.title = element_text(size = 23))
```

Executed code:



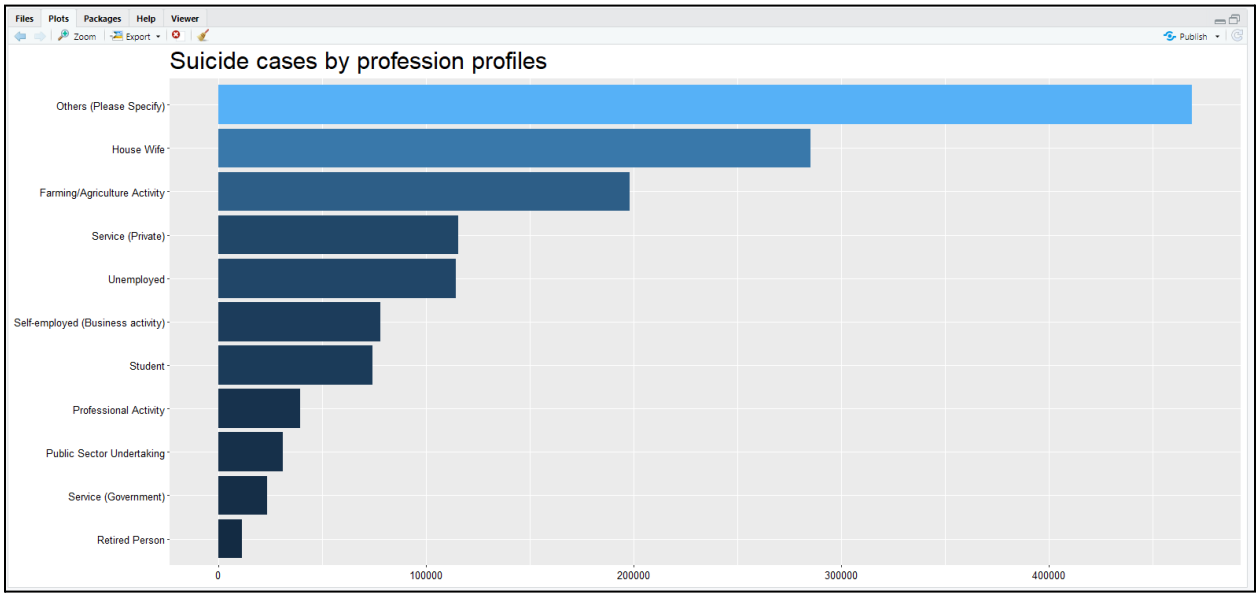

Visualizing suicide cases by professional status:

Sample code:

```
profession_df = suicide_data %>%
  select(Type_code, Type, Total) %>%
  filter(Type_code == "Professional_Profile") %>%
  group_by(Type)%>%
  summarize(Total = sum(Total)) %>%
  arrange (-Total)
profession_df = data.frame(profession_df)
write.csv(profession_df,"profession.csv")
options(scipen = 999)
options(repr.plot.width = 10, repr.plot.height = 15)
plot1 = ggplot(data = profession_df)
plot1 + geom_col(mapping = aes(x = Total, y = reorder(Type, Total), fill = Total)) +
  theme(axis.text = element_text(size = 10, colour = "black")) +
  theme(legend.position = "none") +
  scale_fill_gradient(guide = "colourbar") +
  theme(axis.title = element_blank()) +
```

```
  labs(title = "Suicide cases by profession profiles") +
  theme(plot.title = element_text(size = 23))
```

Executed code:



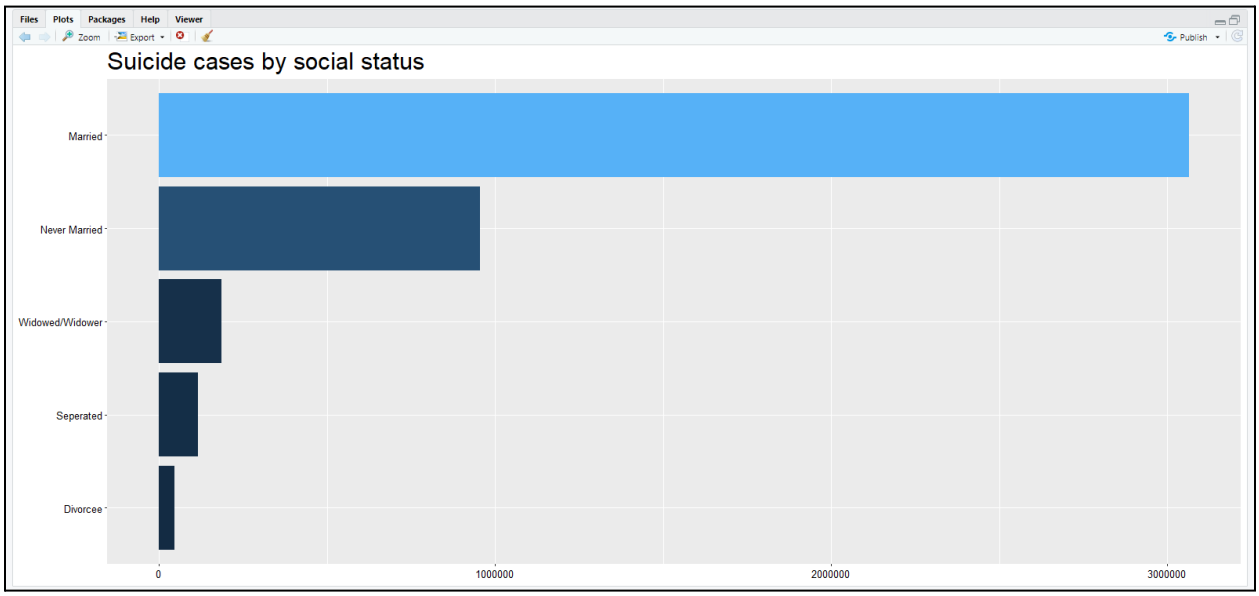

Visualizing suicide cases by social status:

Sample code:

```
social_df = suicide_data %>%
  select(Type_code, Type, Total) %>%
  filter(Type_code == "Social_Status") %>%
  group_by(Type)%>%
  summarize(Total = sum(Total)) %>%
  arrange (-Total)
social_df = data.frame(social_df)
write.csv(social_df,"social.csv")
options(scipen = 999)
options(repr.plot.width = 10, repr.plot.height = 15)
plot1 = ggplot(data = social_df)
plot1 + geom_col(mapping = aes(x = Total, y = reorder(Type, Total), fill = Total)) +
  theme(axis.text = element_text(size = 10, colour = "black")) +
  theme(legend.position = "none") +
  scale_fill_gradient(guide = "colourbar") +
  theme(axis.title = element_blank()) +
```

```
  labs(title = "Suicide cases by social status") +
  theme(plot.title = element_text(size = 23))
```

Executed code:




Visualizing the age wise distribution for the leading causes of suicide cases:

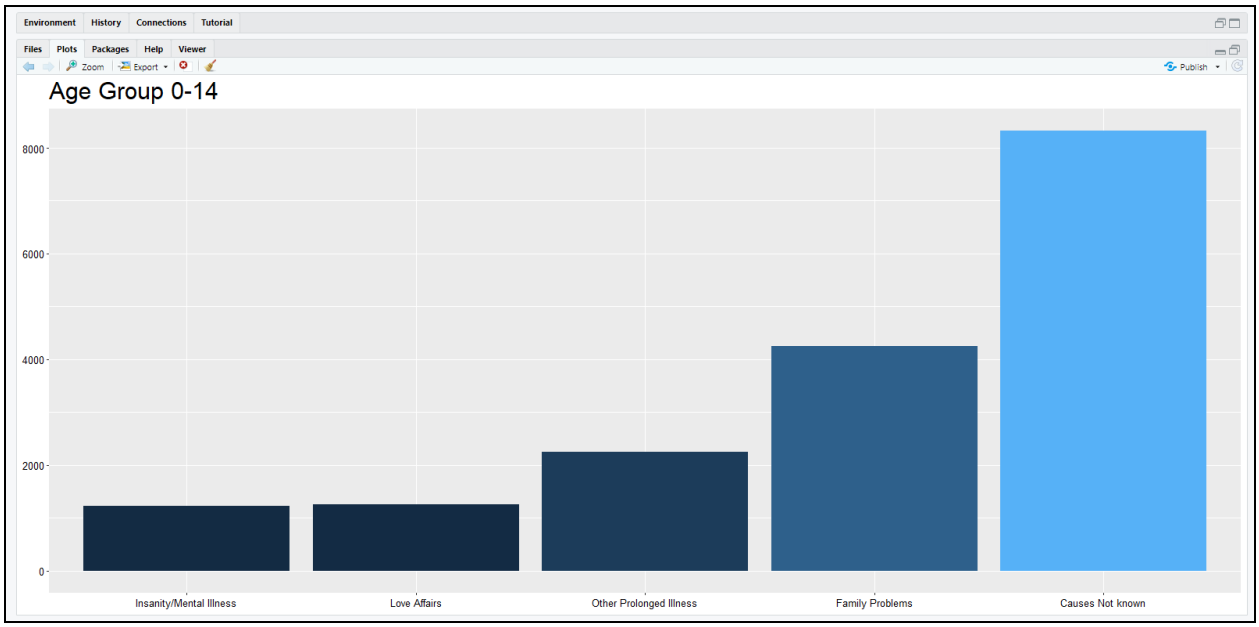
Sample code:

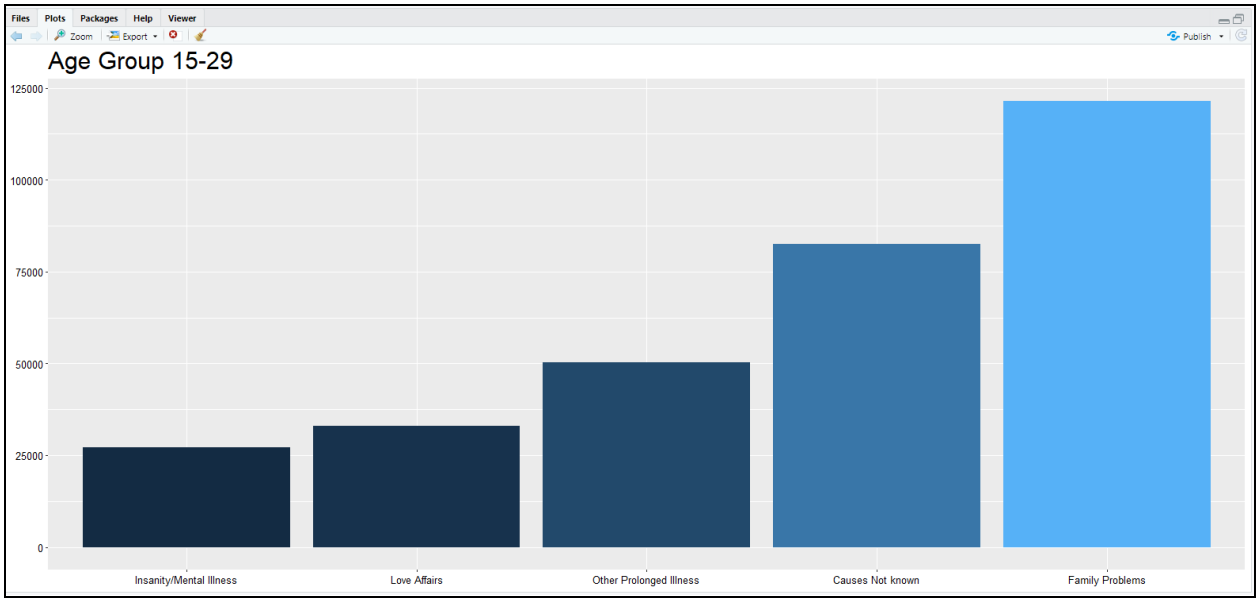
```
Causes_age_wise_df = suicide_data %>%
  select(Type_code, Type, Age_group, Total) %>%
  filter(Type_code == "Causes") %>%
  group_by(Type, Age_group) %>%
  summarize(Total = sum(Total)) %>%
  arrange (Age_group, -Total)
Causes_age_wise_df = data.frame(Causes_age_wise_df)
write.csv(Causes_age_wise_df,"Causes Age Wise.csv")
Leading_causes = c("Family Problems","Causes Not known","Love Affairs","Other Prolonged
Illness","Insanity/Mental Illness")
Top5_df = Causes_age_wise_df %>%
  select(Type, Age_group, Total) %>%
  filter(Type %in% Leading_causes)
Top5_df = data.frame(Top5_df)
write.csv(Top5_df,"Top 5.csv")
#INDIVIDUAL FRAMES FOR EACH AGE GROUP
```

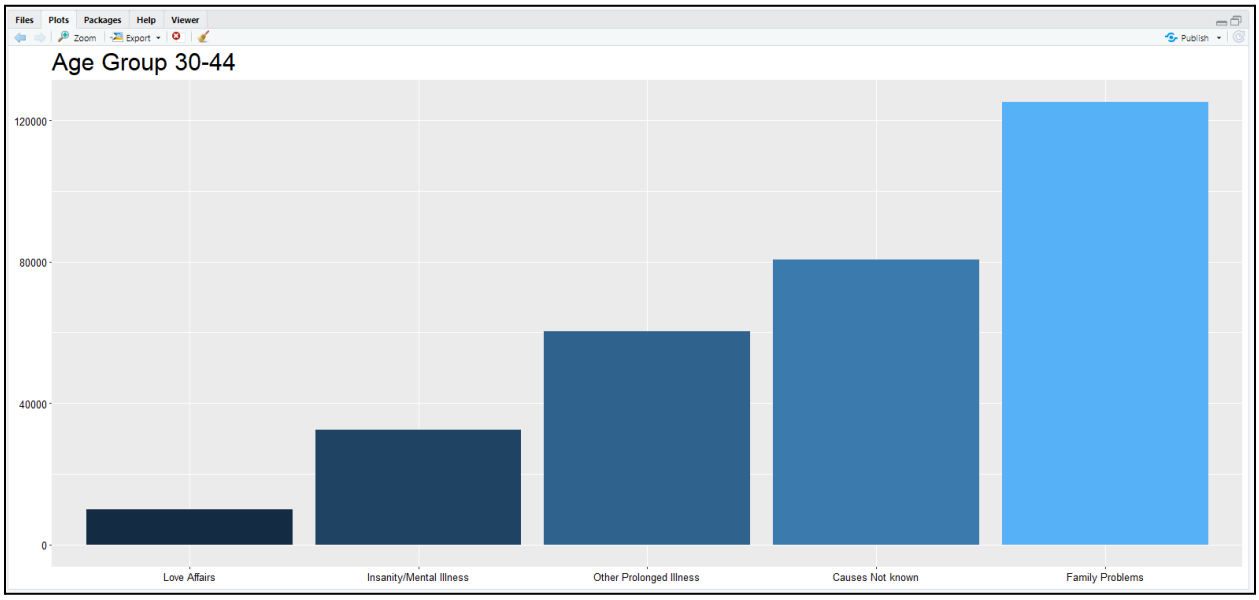
```
Top5_014df = subset(Top5_df,Age_group == "0-14",select = c(Type,Total))
write.csv(Top5_014df,"Top 5 0-14.csv")
Top5_0_14 = table(Type = Top5_014df$Type,Total = Top5_014df$Total)
Top5_0_14
Top5_1529df = subset(Top5_df,Age_group == "15-29",select = c(Type,Total))
write.csv(Top5_1529df,"Top 5 15-29.csv")
Top5_3044df = subset(Top5_df,Age_group == "30-44",select = c(Type,Total))
write.csv(Top5_3044df,"Top 5 30-44.csv")
Top5_4559df = subset(Top5_df,Age_group == "45-59",select = c(Type,Total))
write.csv(Top5_4559df,"Top 5 45-59.csv")
Top5_60df = subset(Top5_df,Age_group == "60+",select = c(Type,Total))
write.csv(Top5_60df,"Top 5 60+.csv")
# INDIVIDUAL GRAPHS FOR AGE-WISE DISTRIBUTION
# AGES 0 TO 14
plot = ggplot(data = Top5_014df)
plot + geom_col(mapping = aes(x = reorder(Type,Total), y = Total, fill = Total)) +
  theme(axis.text = element_text(size = 10, colour = "black")) +
  theme(legend.position = "none") +
  theme(axis.title = element_blank()) +
  labs(title = "Age Group 0-14") +
  theme(plot.title = element_text(size = 23))
# AGES 15 TO 29
plot = ggplot(data = Top5_1529df)
plot + geom_col(mapping = aes(x = reorder(Type,Total), y = Total, fill = Total)) +
  theme(axis.text = element_text(size = 10, colour = "black")) +
  theme(legend.position = "none") +
  theme(axis.title = element_blank()) +
  labs(title = "Age Group 15-29") +
  theme(plot.title = element_text(size = 23))
# AGES 30 T0 44
plot = ggplot(data = Top5_3044df)
plot + geom_col(mapping = aes(x = reorder(Type,Total), y = Total, fill = Total)) +
  theme(axis.text = element_text(size = 10, colour = "black")) +
  theme(legend.position = "none") +
  theme(axis.title = element_blank()) +
  labs(title = "Age Group 30-44") +
  theme(plot.title = element_text(size = 23))
# AGES 45 TO 59
plot = ggplot(data = Top5_4559df)
plot + geom_col(mapping = aes(x = reorder(Type,Total), y = Total, fill = Total)) +
```

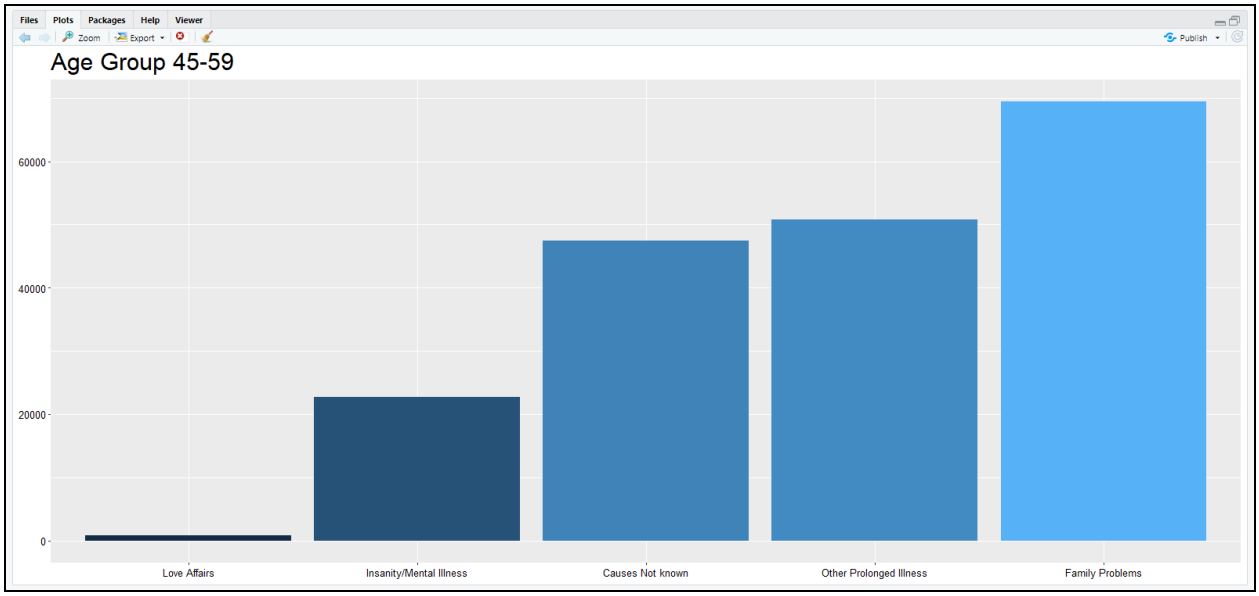
```
  theme(axis.text = element_text(size = 10, colour = "black")) +
  theme(legend.position = "none") +
  theme(axis.title = element_blank()) +
  labs(title = "Age Group 45-59") +
  theme(plot.title = element_text(size = 23))
# AGES 60+
plot = ggplot(data = Top5_60df)
plot + geom_col(mapping = aes(x = reorder(Type,Total), y = Total, fill = Total)) +
  theme(axis.text = element_text(size = 10, colour = "black")) +
  theme(legend.position = "none") +
  theme(axis.title = element_blank()) +
  labs(title = "Age Group 60+") +
  theme(plot.title = element_text(size = 23))
#GROUP BAR PLOT REPRESENTATION FOR ALL AGES
ggplot(aes(x = Type,y = Total,group = Age_group, fill = Age_group),data = Top5_df)+
  geom_bar(position = "dodge", stat = "identity")
```

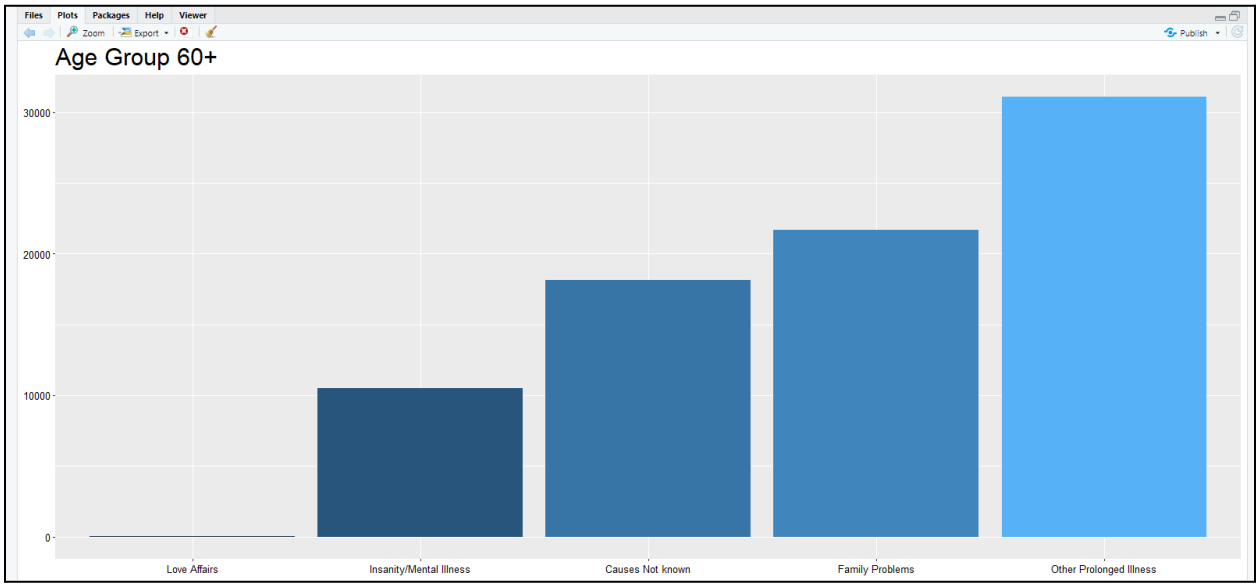
Executed code:

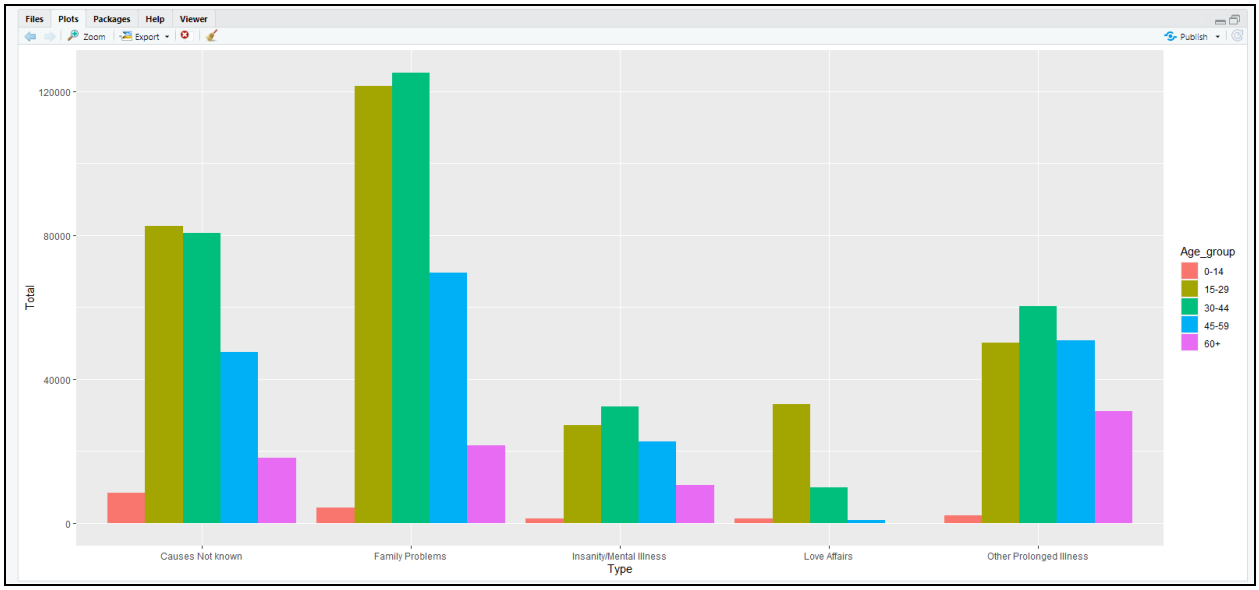



Files
Plots
Packages
Help
Viewer
Zoom
Export
Publish
Age Group 15-29
125000
100000
75000
50000
25000
0
Insanity/Mental Illness
Love Affairs
Other Prolonged Illness
Causes Not known
Family Problems


Files
Plots
Packages
Help
Viewer
Zoom
Export
Publish
Age Group 30-44
120000
80000
40000
0
Love Affairs
Insanity/Mental Illness
Other Prolonged Illness
Causes Not known
Family Problems


Files
Plots
Packages
Help
Viewer
Zoom
Export
Publish
Age Group 45-59
60000
40000
20000
0
Love Affairs
Insanity/Mental Illness
Causes Not known
Other Prolonged Illness
Family Problems


Files
Plots
Packages
Help
Viewer
Zoom
Export
Publish
Age Group 60+
30000
20000
10000
0
Love Affairs
Insanity/Mental Illness
Causes Not known
Family Problems
Other Prolonged Illness




**LINEAR REGRESSION:**

Drawing a relation between the total deaths and years for the timeline 2001-2012:

Sample code:

```
relation = lm(Total~Year,data = TotalTally)
summary(relation)
```

Executed code:

```
> relation = lm(Total~Year,data = TotalTally)
> summary(relation)

Call:
lm(formula = Total ~ Year, data = TotalTally)

Residuals:
   Min     1Q Median     3Q    Max
-31243  -9847  -2390  13177  38647

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -46701885    3605386  -12.95 1.42e-07 ***
Year            23818       1797   13.26 1.14e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 21490 on 10 degrees of freedom
Multiple R-squared:  0.9462,    Adjusted R-squared:  0.9408
F-statistic: 175.7 on 1 and 10 DF,  p-value: 1.14e-07
```

Drawing a relation for the total deaths and years for the timeline 2001-2006:

Sample code:

```
TotalTally_sub = subset(TotalTally,Year<2007,select = c(Year,Total))
TotalTally_sub
sub_relation = lm(Total~Year,data = TotalTally_sub)
summary(sub_relation)
```

Executed code:

```
> TotalTally_sub = subset(TotalTally,Year<2007,select = c(Year,Total))
> TotalTally_sub
  Year   Total
1 2001  976464
2 2002  993648
3 2003  997622
4 2004 1023137
5 2005 1025201
6 2006 1062991
```

```
> sub_relation = lm(Total~Year,data = TotalTally_sub)
> summary(sub_relation)

Call:
lm(formula = Total ~ Year, data = TotalTally_sub)

Residuals:
     1      2      3      4      5      6
  2773   4163  -7658   2063 -11668  10327

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -30631189    4354787  -7.034  0.00215 **
Year            15794       2174   7.267  0.00191 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9093 on 4 degrees of freedom
Multiple R-squared:  0.9296,    Adjusted R-squared:  0.912
F-statistic:  52.8 on 1 and 4 DF,  p-value: 0.001905
```

Predicting the total number of deaths for the timeline 2007-2012:

Sample code:

```
predict_years = data.frame(Year = c(2007,2008,2009,2010,2011,2012))
predictions <- predict(sub_relation,newdata = predict_years)
predictions
years = seq(2001,2012,1)
prediction_table = data.frame(Year = years,Total = c(NA,NA,NA,NA,NA,NA,predictions))
prediction_table
TotalTally
```

Executed code:

```
> predict_years = data.frame(Year = c(2007,2008,2009,2010,2011,2012))
> predictions <- predict(sub_relation,newdata = predict_years)
> predictions
      1       2       3       4       5       6
1068458 1084253 1100047 1115842 1131636 1147431
> years = seq(2001,2012,1)
> prediction_table = data.frame(Year = years,Total = c(NA,NA,NA,NA,NA,NA,predictions))
> prediction_table
   Year   Total
1  2001      NA
2  2002      NA
3  2003      NA
4  2004      NA
5  2005      NA
6  2006      NA
7  2007 1068458
8  2008 1084253
9  2009 1100047
10 2010 1115842
11 2011 1131636
12 2012 1147431
```

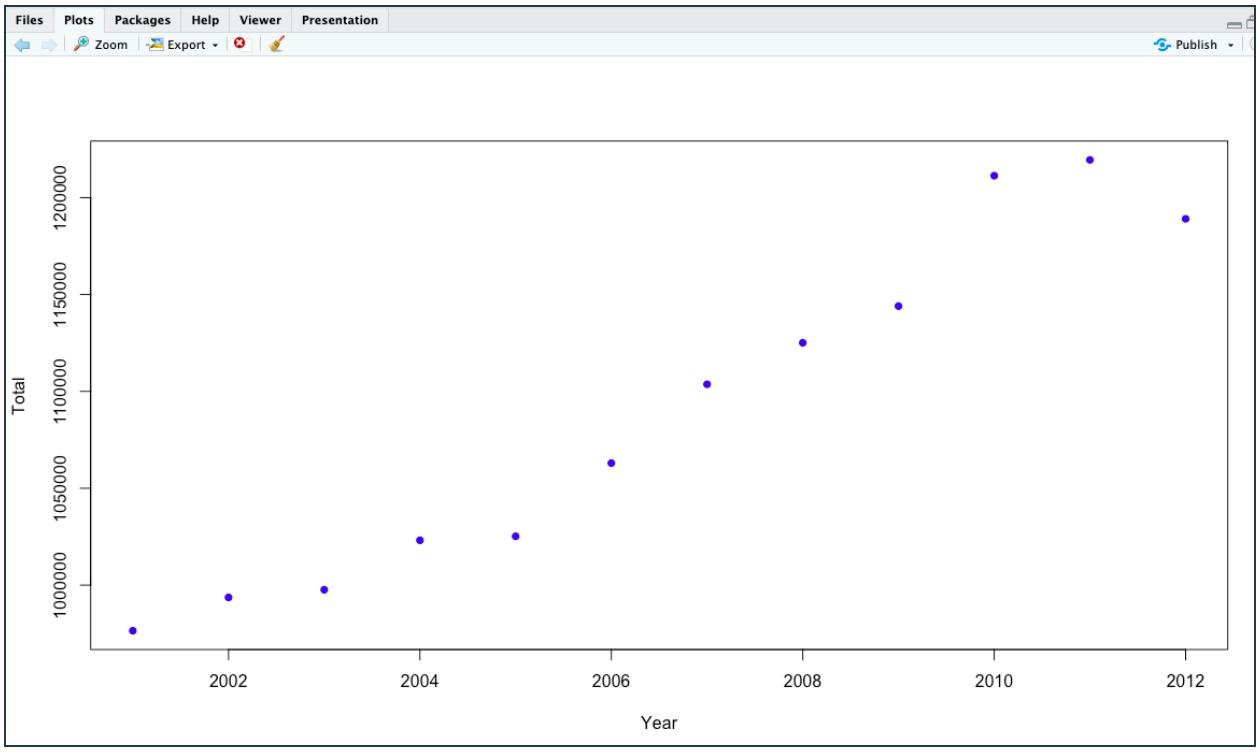
```
> TotalTally
   Year   Total
1  2001  976464
2  2002  993648
3  2003  997622
4  2004 1023137
5  2005 1025201
6  2006 1062991
7  2007 1103667
8  2008 1125082
9  2009 1144033
10 2010 1211322
11 2011 1219499
12 2012 1189068
```

Plotting the original dataset:

Sample code:

plot(TotalTally, pch = 16, col = "blue")

Executed code:

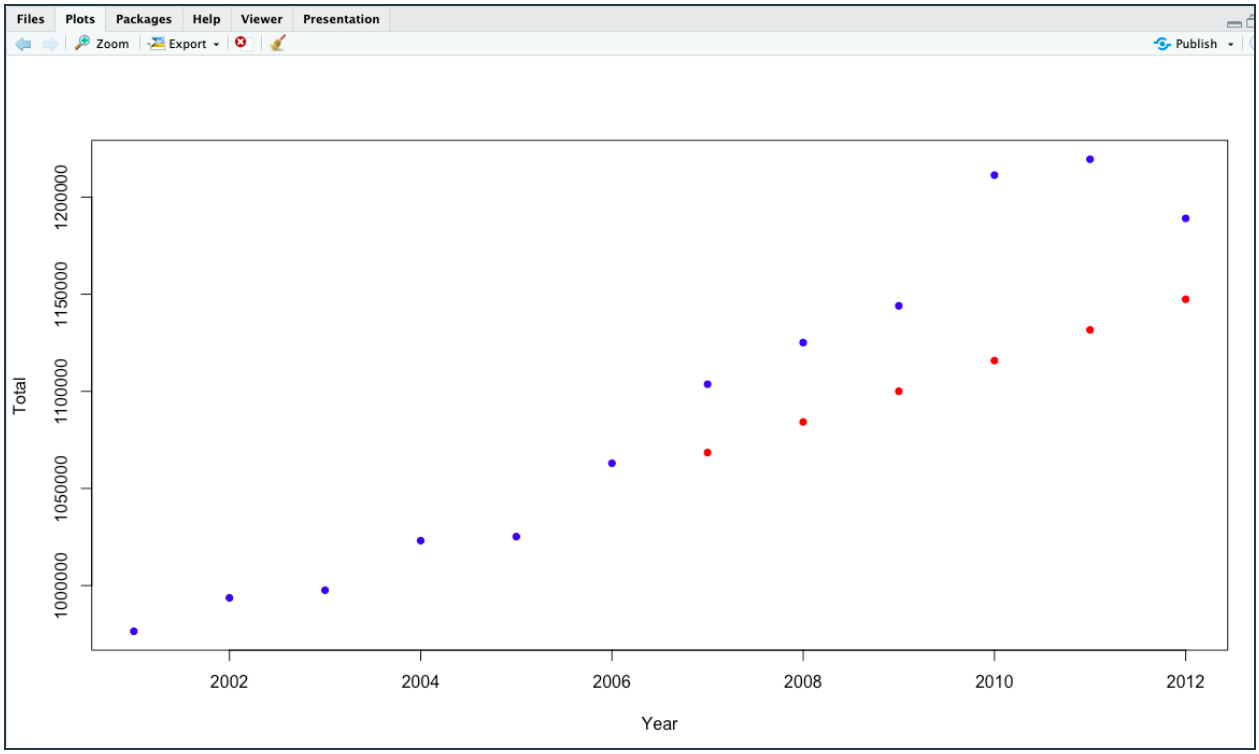



Plotting the predictions and comparing them with original values:

Sample code:

points(prediction_table,col = "red",pch =16)

Executed code:

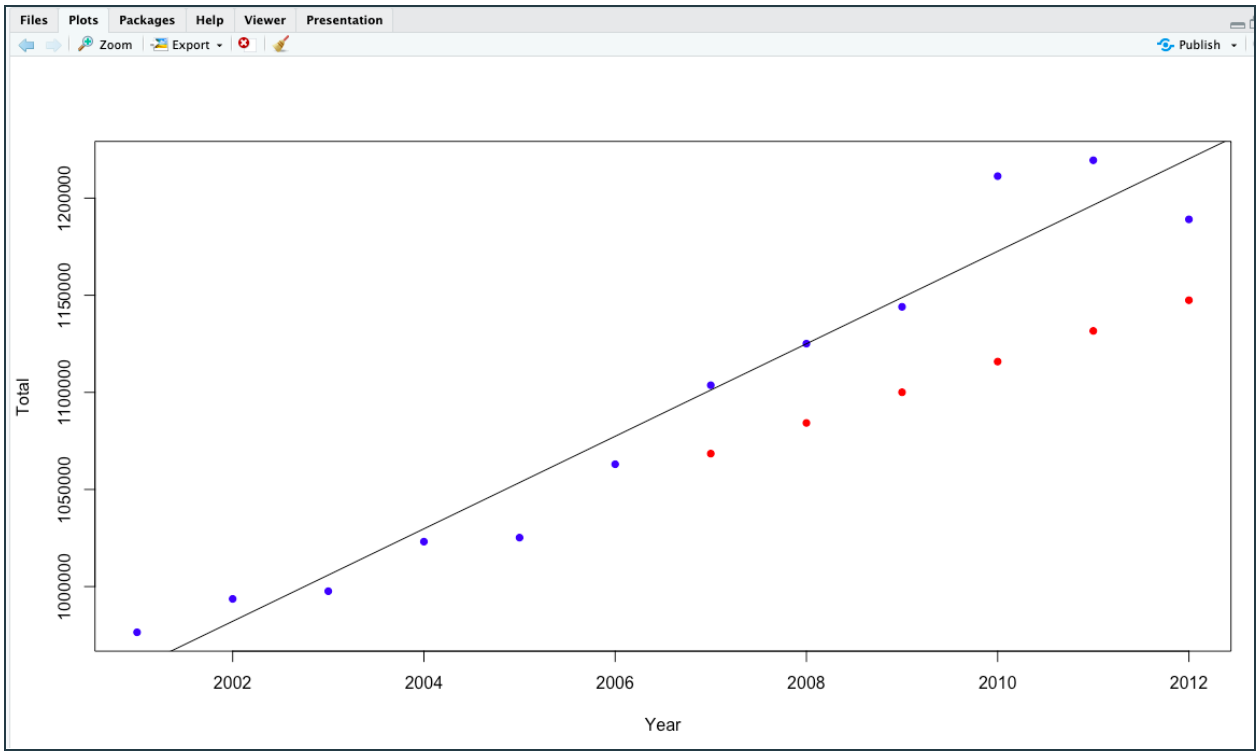



Plotting the regression model:

Sample code:

abline(relation)

Executed code:




**CONCLUSIONS:**

Through this project, we were able to visualize and interpret various aspects of Indian suicide cases with the help of a variety of tools in R. This gave us a lot of insight into the complexities of the issue and broadened our scope of information on the subject. We gained a lot of experience on coding in the R language and interacting with the RStudio interface. We also gained the ability to apply the concepts of Inference and Interpretation on real world problems. Working on this project also enhanced our time management, communication and group coordination skills.

**REFERENCES:**


- https://www.tutorialspoint.com/r/r_linear_regression.htm
- Kosuke Imai, Quantitative Social Science - Princeton University Press, 2018
- https://www.w3schools.com/r/default.asp
- https://www.tutorialspoint.com/r/r_multiple_regression.htm
- https://www.geeksforgeeks.org/simple-linear-regression-using-r/
- https://www.tutorialspoint.com/r/r_bar_charts.htm
- https://www.tutorialspoint.com/r/r_line_graphs.htm