

SUICIDE DATASET

Vijaya Suresh

2023-03-20

#INSTALLING REQUIRED LIBRARIES

```
install.packages("dplyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'  
## (as 'lib' is unspecified)
```

```
install.packages("rmarkdown")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'  
## (as 'lib' is unspecified)
```

```
install.packages("purrr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'  
## (as 'lib' is unspecified)
```

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'  
## (as 'lib' is unspecified)
```

```
install.packages("reshape2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'  
## (as 'lib' is unspecified)
```

```
library(readxl)  
library(ggplot2)  
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(rmarkdown)  
library(purrr)  
library(tidyverse)
```

```
## — Attaching core tidyverse packages ————— tidyverse 2.0.0 —
```

```
## ✓ forcats 1.0.0      ✓ stringr 1.5.0  
## ✓ lubridate 1.9.2    ✓ tibble 3.2.0  
## ✓ readr 2.1.4       ✓ tidyr 1.3.0
```

```
## — Conflicts ————— tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag() masks stats::lag()
## ⓘ Use the ]8;;http://conflicted.r-lib.org/conflicted-package]8;; to force all
conflicts to become errors
```

```
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
##
## The following object is masked from 'package:tidyr':
##
## smiths
```

#IMPORTING THE DATASET

```
data <- read_excel("csv vijaya.xlsx")
print(data)
```

```
## # A tibble: 93 × 5
##   Category `State/UT/City` `Number of Suicides - 2020` Number of Su...1 Perce...2
##   <chr>      <chr>                <dbl>          <dbl>    <dbl>
## 1 State      Andhra Pradesh                7043            8067    14.5
## 2 State      Arunachal Pradesh                 160             160      0
## 3 State      Assam                      3243            3262    0.6
## 4 State      Bihar                       809             827    2.2
## 5 State      Chhattisgarh                7710            7828    1.5
## 6 State      Goa                          308             321    4.2
## 7 State      Gujarat                    8050            8789    9.2
## 8 State      Haryana                     4001            3692   -7.7
## 9 State      Himachal Pradesh              857             889    3.7
## 10 State     Jharkhand                    2145            1825  -14.9
## # ... with 83 more rows, and abbreviated variable names
## #   1`Number of Suicides - 2021`, 2`Percentage Variation in 20201over 2020`
```

#BASIC INSIGHTS

```
glimpse(data)
```

```
## Rows: 93
## Columns: 5
## $ Category                <chr> "State", "State", "State", "S...
## $ `State/UT/City`         <chr> "Andhra Pradesh", "Arunachal ...
## $ `Number of Suicides - 2020` <dbl> 7043, 160, 3243, 809, 7710, 3...
## $ `Number of Suicides - 2021` <dbl> 8067, 160, 3262, 827, 7828, 3...
## $ `Percentage Variation in 20201over 2020` <dbl> 14.5, 0.0, 0.6, 2.2, 1.5, 4.2...
```

```
distinct(data)
```

```
## # A tibble: 93 × 5
##   Category `State/UT/City` `Number of Suicides - 2020` Number of Su...1 Perce...2
##   <chr>      <chr>                <dbl>          <dbl>    <dbl>
## 1 State      Andhra Pradesh                7043            8067    14.5
## 2 State      Arunachal Pradesh                 160             160      0
## 3 State      Assam                      3243            3262    0.6
## 4 State      Bihar                       809             827    2.2
## 5 State      Chhattisgarh                7710            7828    1.5
## 6 State      Goa                          308             321    4.2
```

```
## 7 State Gujarat 8050 8789 9.2
## 8 State Haryana 4001 3692 -7.7
## 9 State Himachal Pradesh 857 889 3.7
## 10 State Jharkhand 2145 1825 -14.9
## # ... with 83 more rows, and abbreviated variable names
## # 1`Number of Suicides - 2021`, 2`Percentage Variation in 20201over 2020`
```

#CHECKING NULL VALUES

```
data%>% map(~sum(is.na(.)))
```

```
## $Category
## [1] 3
##
## $`State/UT/City`
## [1] 0
##
## $`Number of Suicides - 2020`
## [1] 0
##
## $`Number of Suicides - 2021`
## [1] 0
##
## $`Percentage Variation in 20201over 2020`
## [1] 0
```

#DELETING NULL VALUE ROWS

```
data_1 = data[-29,]
data_1 = data_1[-38,]
data_1 = data_1[-91,]
```

```
summary(data_1)
```

```
## Category State/UT/City Number of Suicides - 2020
## Length:90 Length:90 Min. : 2.0
## Class :character Class :character 1st Qu.: 163.2
## Mode :character Mode :character Median : 332.5
## Mean : 2013.6
## 3rd Qu.: 1197.2
## Max. :19909.0
## Number of Suicides - 2021 Percentage Variation in 20201over 2020
## Min. : 1.0 Min. : -50.400
## 1st Qu.: 174.2 1st Qu.: -8.150
## Median : 378.5 Median : 4.300
## Mean : 2155.3 Mean : 7.686
## 3rd Qu.: 1341.8 3rd Qu.: 12.250
## Max. :22207.0 Max. :200.000
```

#RENAMING COLUMN NAMES

```
colnames(data_1)[2]='Location'
colnames(data_1)[3]="Suicide_2020"
colnames(data_1)[4]="Suicide_2021"
colnames(data_1)[5]="Variation"
```

#MUTATE A COLUMN VALUE

```
data_1 = data_1 %>%
  select(Category,Location,Suicide_2020,Suicide_2021,Variation)%>%
```

```

mutate(Variation =Suicide_2021- Suicide_2020 )

head(data_1)

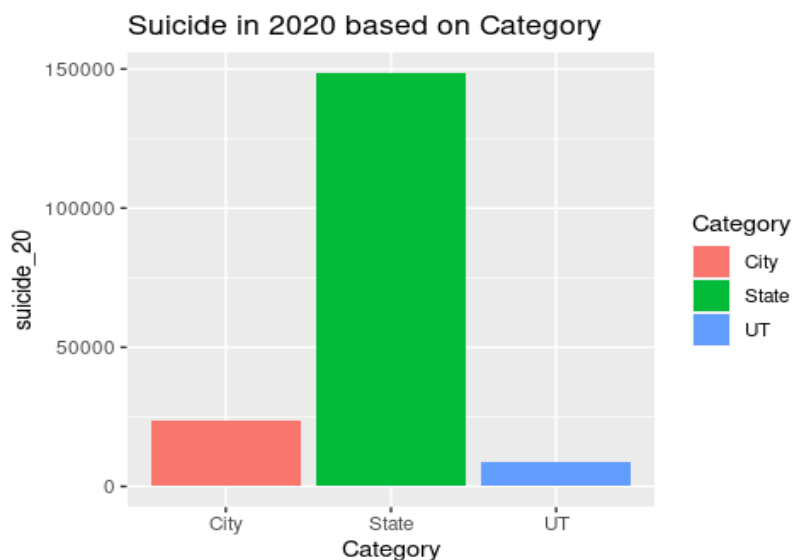
## # A tibble: 6 × 5
##   Category Location      Suicide_2020 Suicide_2021 Variation
##   <chr>    <chr>          <dbl>         <dbl>      <dbl>
## 1 State    Andhra Pradesh      7043         8067      1024
## 2 State    Arunachal Pradesh    160          160         0
## 3 State    Assam                3243         3262        19
## 4 State    Bihar                809          827         18
## 5 State    Chhattisgarh        7710         7828       118
## 6 State    Goa                  308          321         13

#EXPLORATORY DATA ANALYSIS
#CATEGORY WISE ANALYSIS
x= data_1%>%group_by(Category)%>%
  summarise(suicide_20 = sum(Suicide_2020))
x

## # A tibble: 3 × 2
##   Category suicide_20
##   <chr>         <dbl>
## 1 City          23855
## 2 State        148737
## 3 UT           8630

ggplot(x,aes(Category,suicide_20,fill=Category))+
  geom_bar(position="dodge",stat='identity')+
  labs(title = "Suicide in 2020 based on Category")

```



```

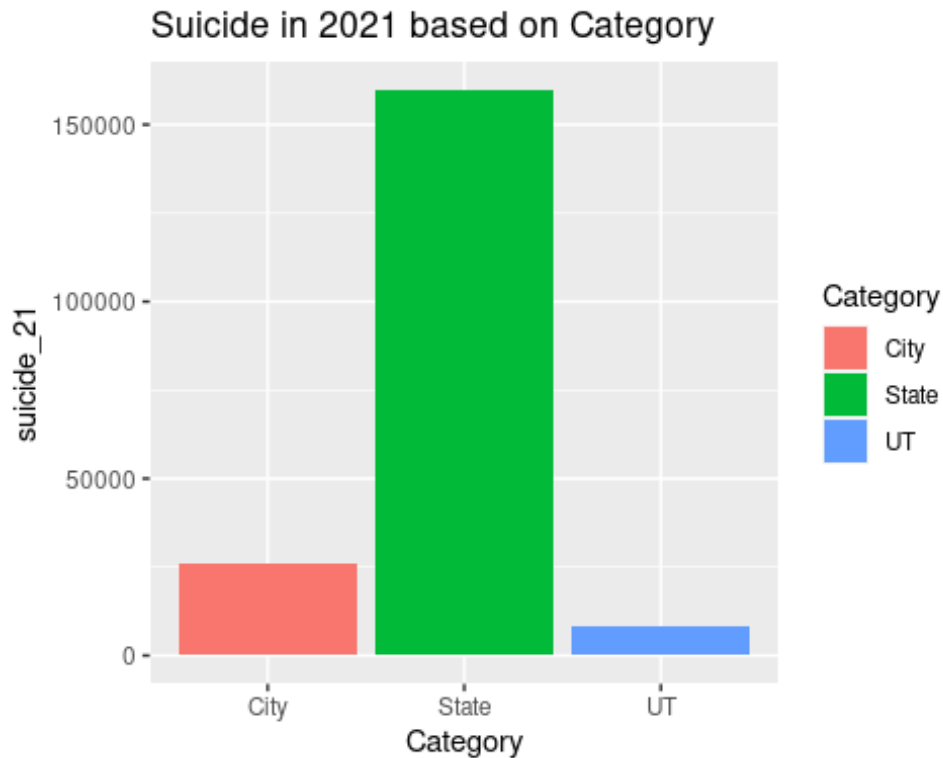
y= data_1%>%group_by(Category)%>%
  summarise(suicide_21 = sum(Suicide_2021))
y

## # A tibble: 3 × 2
##   Category suicide_21
##   <chr>         <dbl>

```

```
## 1 City      25891
## 2 State    159980
## 3 UT       8106

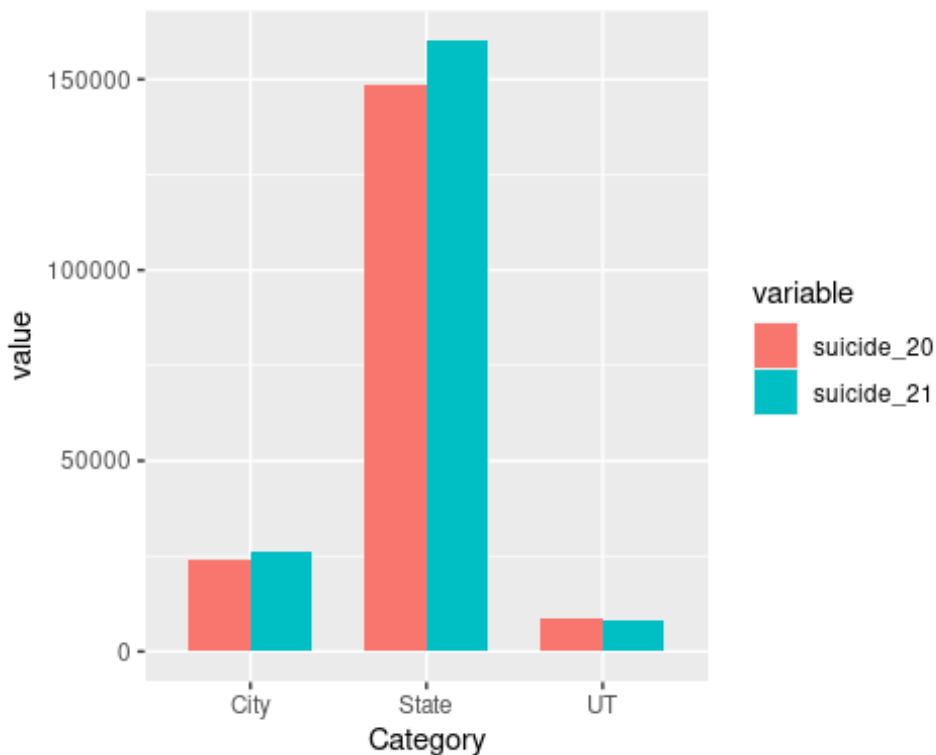
ggplot(y,aes(Category,suicide_21,fill=Category))+
  geom_bar(position="dodge",stat='identity')+
  labs(title = "Suicide in 2021 based on Category")
```



```
x_y =inner_join(x,y)
## Joining with `by = join_by(Category)`
x_y
## # A tibble: 3 x 3
##   Category suicide_20 suicide_21
##   <chr>      <dbl>      <dbl>
## 1 City      23855      25891
## 2 State    148737     159980
## 3 UT       8630       8106

x_y = melt(x_y,id=c('Category'))

ggplot(x_y)+
  geom_bar(aes(x=Category,y=value,fill=variable),
    stat="identity", position = "dodge", width = 0.7)
```



#STATE WISE ANALYSIS

```
state = filter(data_1,Category == 'State')
print(state)
```

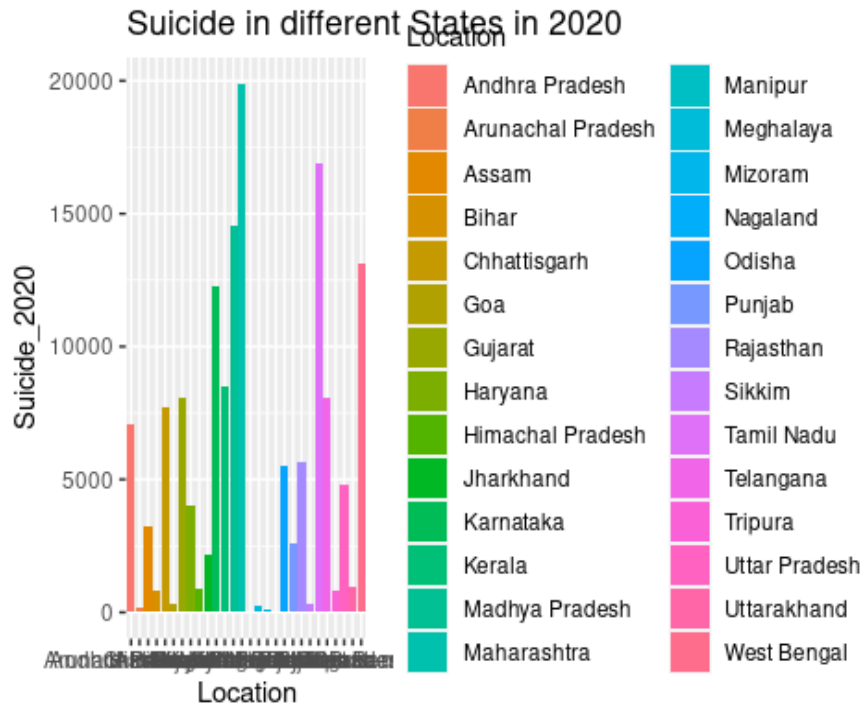
```
## # A tibble: 28 × 5
##   Category Location      Suicide_2020 Suicide_2021 Variation
##   <chr>    <chr>          <dbl>         <dbl>    <dbl>
## 1 State    Andhra Pradesh      7043         8067      1024
## 2 State    Arunachal Pradesh    160          160         0
## 3 State    Assam               3243         3262        19
## 4 State    Bihar                809          827         18
## 5 State    Chhattisgarh        7710         7828        118
## 6 State    Goa                  308          321         13
## 7 State    Gujarat             8050         8789        739
## 8 State    Haryana             4001         3692       -309
## 9 State    Himachal Pradesh     857          889         32
## 10 State   Jharkhand           2145         1825       -320
## # ... with 18 more rows
```

```
summary(state)
```

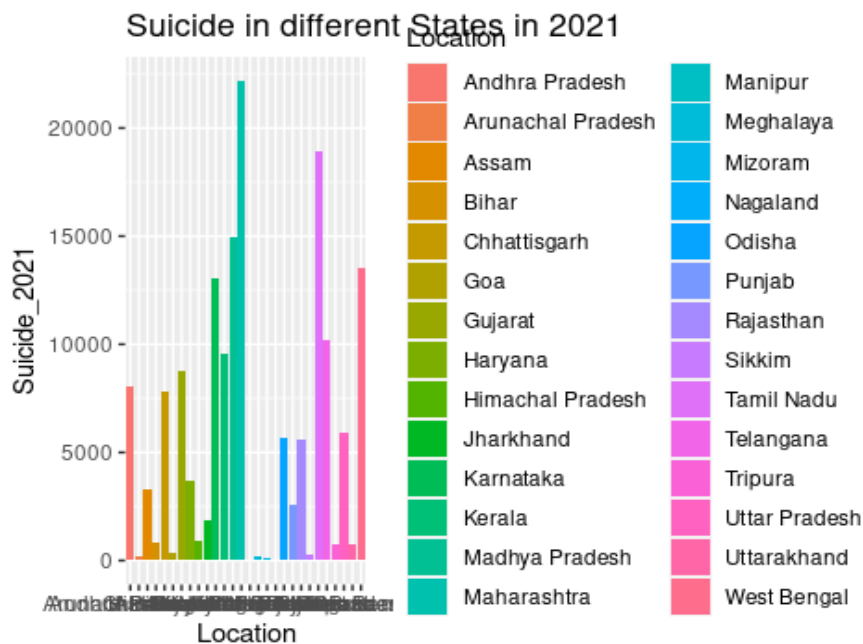
```
##   Category      Location      Suicide_2020      Suicide_2021
## Length:28      Length:28      Min.   : 44.0      Min.   : 43
## Class :character Class :character 1st Qu.: 683.8     1st Qu.: 618
## Mode  :character Mode  :character Median : 3622.0     Median : 3477
##                                     Mean  : 5312.0     Mean   : 5714
##                                     3rd Qu.: 8052.0    3rd Qu.: 8979
##                                     Max.   :19909.0    Max.   :22207
## Variation
## Min.   : -320.00
## 1st Qu.: -10.75
## Median: 18.50
## Mean: 401.54
```

```
## 3rd Qu.: 753.50
## Max. :2298.00
```

```
ggplot(state,aes(Location,Suicide_2020,fill=Location))+
  geom_bar(position="dodge",stat='identity')+
  labs(title = "Suicide in different States in 2020")
```

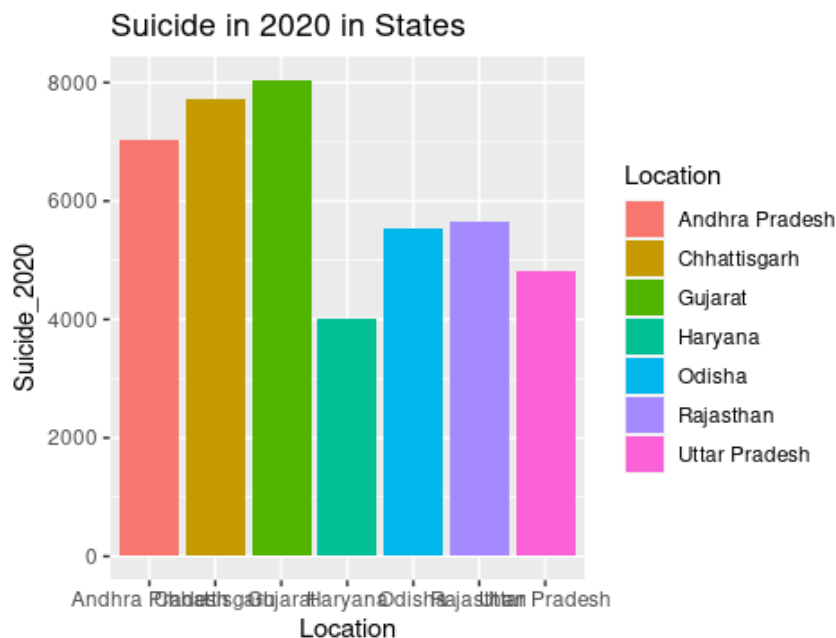


```
ggplot(state,aes(Location,Suicide_2021,fill=Location))+
  geom_bar(position="dodge",stat='identity')+
  labs(title = "Suicide in different States in 2021")
```



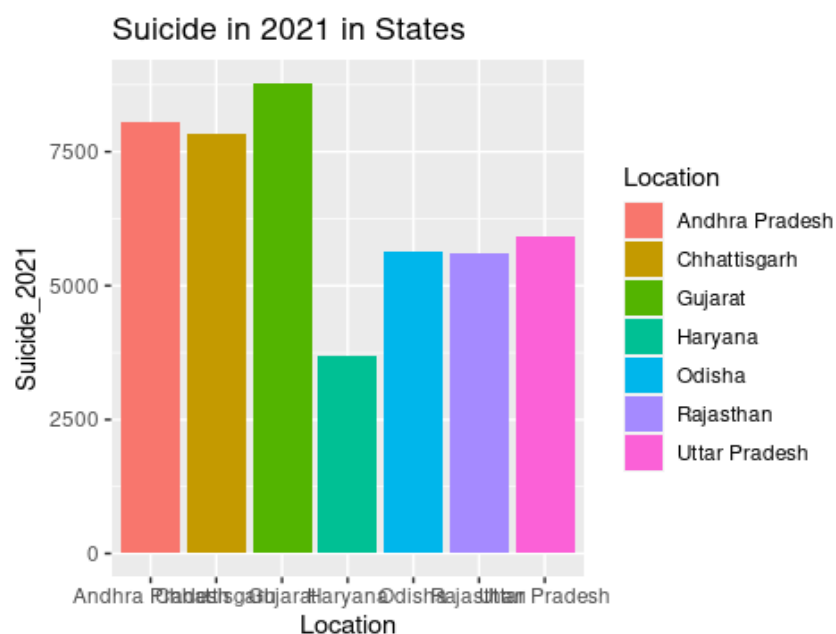
```
state_2020 = filter(state,Suicide_2020 >=3600 & Suicide_2020<=8052)
```

```
ggplot(state_2020,aes(Location,Suicide_2020,fill=Location))+
  geom_bar(position="dodge",stat='identity')+
  labs(title = "Suicide in 2020 in States")
```



```
state_2021 = filter(state,Suicide_2021 >=3400 & Suicide_2021<=8980)
```

```
ggplot(state_2021,aes(Location,Suicide_2021,fill=Location))+
  geom_bar(position="dodge",stat='identity')+
  labs(title = "Suicide in 2021 in States")
```




```

state_1=inner_join(state_2020,state_2021)

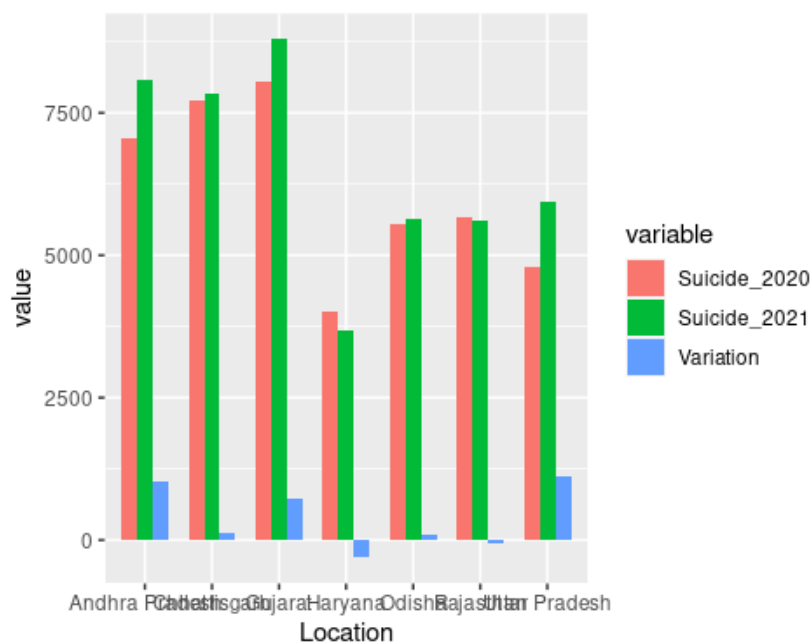
## Joining with `by = join_by(Category, Location, Suicide_2020, Suicide_2021,
## Variation)`

state_1 = state_1[,-1]

state_20_21 = melt(state_1,id=c('Location'))

ggplot(state_20_21)+
  geom_bar(aes(x=Location,y=value,fill=variable),
           stat="identity", position = "dodge", width = 0.7)

```



#CITY WISE ANALYSIS

```

city = filter(data_1,Category == 'City')
print(city)

```

```

## # A tibble: 53 × 5
##   Category Location      Suicide_2020 Suicide_2021 Variation
##   <chr>    <chr>          <dbl>         <dbl>    <dbl>
## 1 City    Agra             115           99      -16
## 2 City    Ahmedabad       871          991     120
## 3 City    Allahabad        40           45        5
## 4 City    Amritsar         97           69     -28
## 5 City    Asansol          329          477    148
## 6 City    Aurangabad       270          297     27
## 7 City    Bengaluru       2196         2292     96
## 8 City    Bhopal           416          566    150
## 9 City    Chandigarh (City) 128          120     -8
## 10 City   Chennai        2430         2699    269
## # ... with 43 more rows

```

```
summary(city)
```

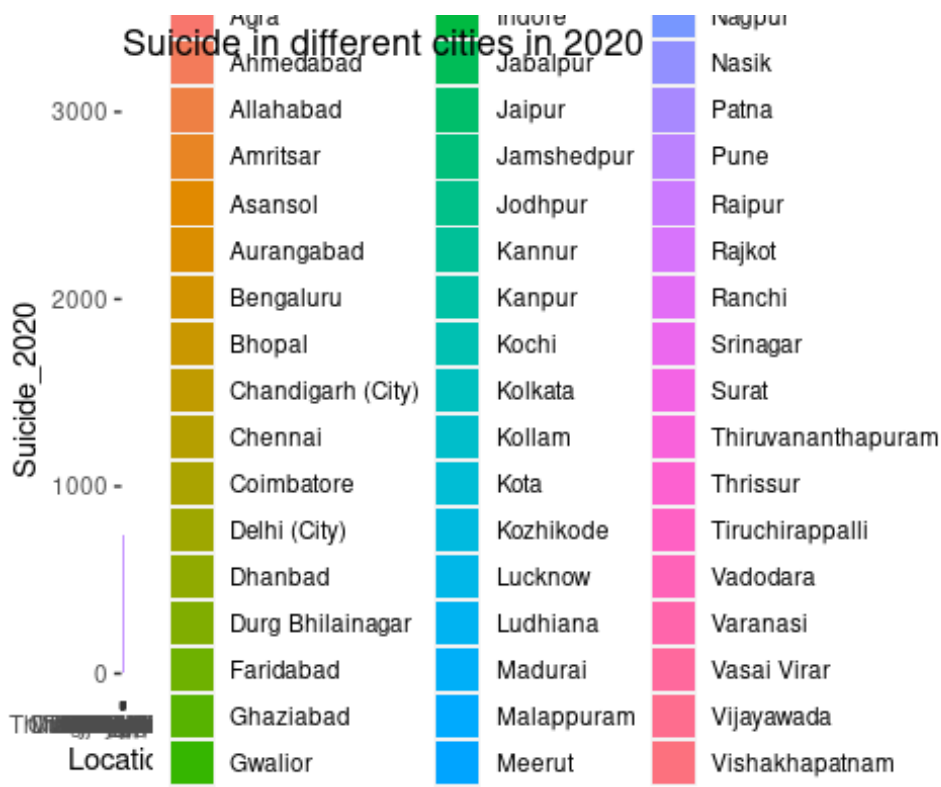
```

##   Category      Location      Suicide_2020      Suicide_2021
## Length:53      Length:53      Min.   : 6.0      Min.   : 18.0
## Class :character Class :character 1st Qu.: 147.0   1st Qu.: 160.0

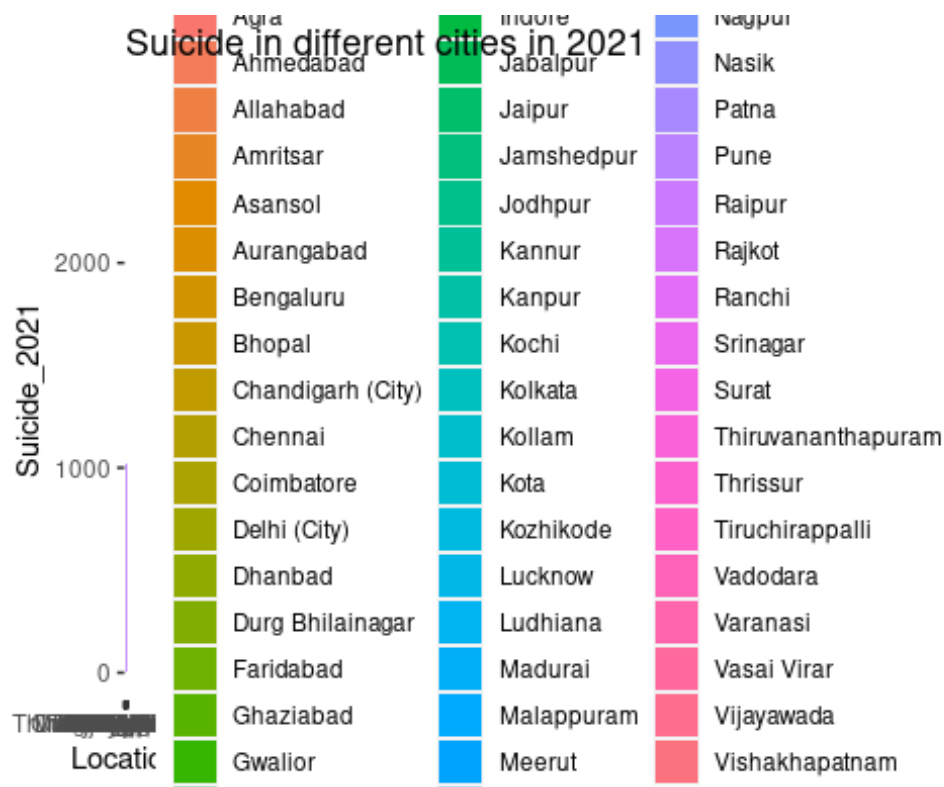
```

```
## Mode :character Mode :character Median : 312.0 Median : 338.0
## Mean : 450.1 Mean : 488.5
## 3rd Qu.: 417.0 3rd Qu.: 477.0
## Max. :3025.0 Max. :2760.0
## Variation
## Min. : -265.00
## 1st Qu.: -11.00
## Median : 20.00
## Mean : 38.42
## 3rd Qu.: 88.00
## Max. : 282.00
```

```
ggplot(city,aes(Location,Suicide_2020,fill=Location))+
  geom_bar(position="dodge",stat='identity')+
  labs(title = "Suicide in different cities in 2020")
```

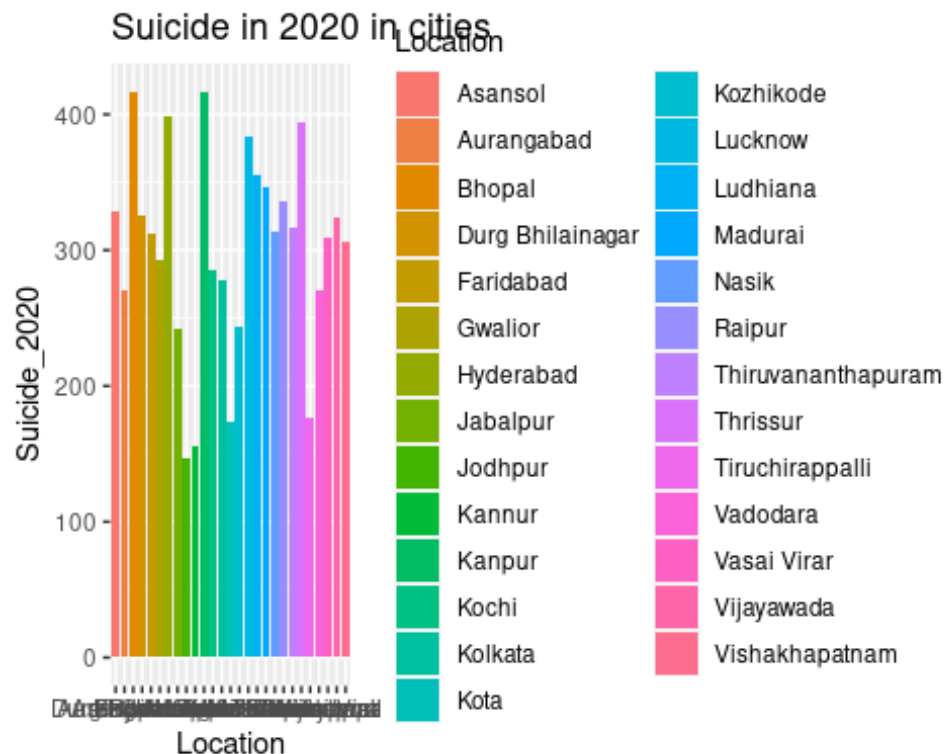


```
ggplot(city,aes(Location,Suicide_2021,fill=Location))+
  geom_bar(position="dodge",stat='identity')+
  labs(title = "Suicide in different cities in 2021")
```



```
city_2020 = filter(city,Suicide_2020 >=147 & Suicide_2020<=417)
```

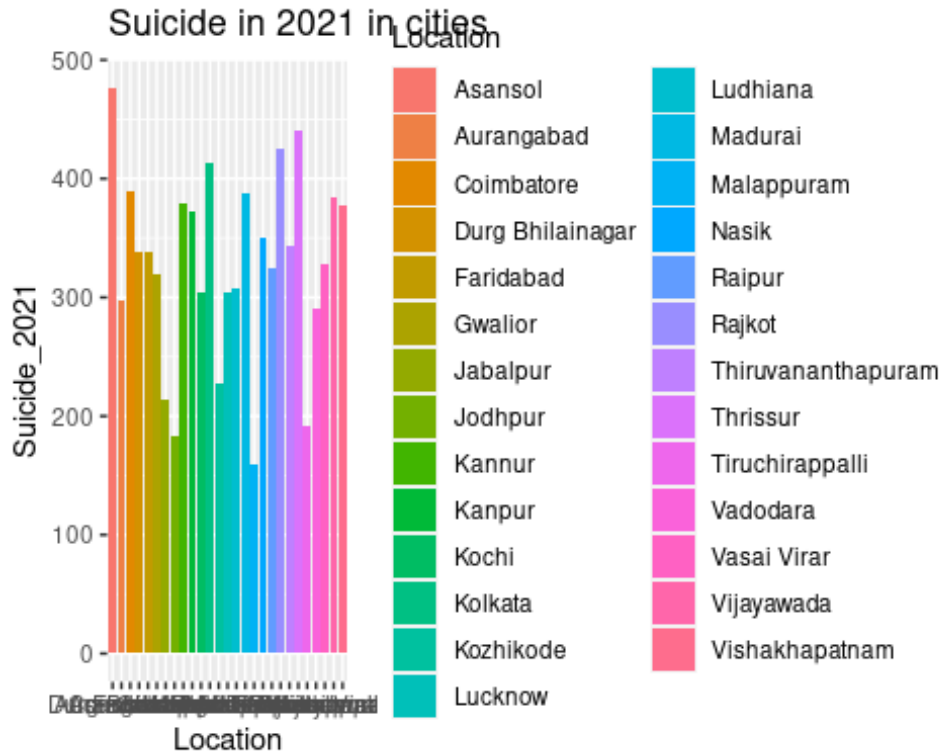
```
ggplot(city_2020,aes(Location,Suicide_2020,fill=Location))+
  geom_bar(position="dodge",stat='identity')+
  labs(title = "Suicide in 2020 in cities")
```



```
city_2021 = filter(city,Suicide_2021 >=160 & Suicide_2021<=477)
```

```
ggplot(city_2021,aes(Location,Suicide_2021,fill=Location))+
```

```
geom_bar(position="dodge",stat='identity')+
labs(title = "Suicide in 2021 in cities")
```



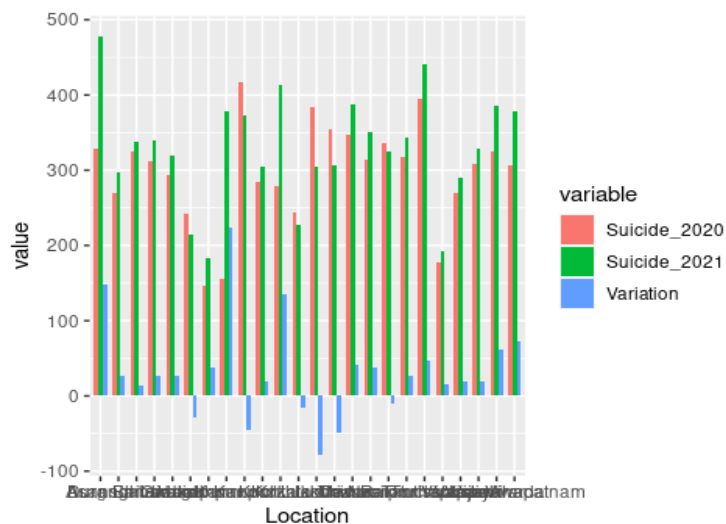
```
city_1=inner_join(city_2020,city_2021)

## Joining with `by = join_by(Category, Location, Suicide_2020, Suicide_2021,
## Variation)`

city_1 = city_1[,-1]

city_20_21 = melt(city_1,id=c('Location'))

ggplot(city_20_21)+
  geom_bar(aes(x=Location,y=value,fill=variable),
    stat="identity", position = "dodge", width = 0.7)
```



#UNION TERRITORY WISE ANALYSIS

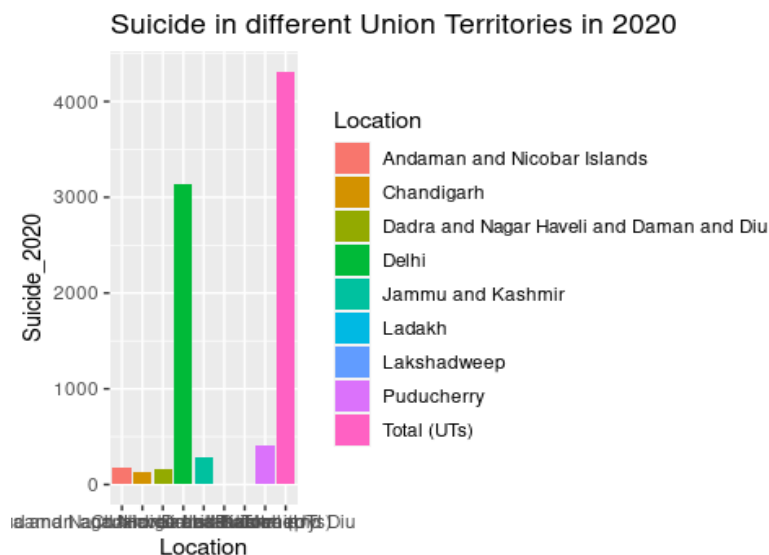
```
ut = filter(data_1,Category == 'UT')
print(ut)
```

```
## # A tibble: 9 × 5
##   Category Location                               Suicide_2020 Suici...1 Varia...2
##   <chr>    <chr>                                <dbl>    <dbl>    <dbl>
## 1 UT      Andaman and Nicobar Islands              180      159     -21
## 2 UT      Chandigarh                             128      120      -8
## 3 UT      Dadra and Nagar Haveli and Daman and Diu  156      171      15
## 4 UT      Delhi                                3142     2840    -302
## 5 UT      Jammu and Kashmir                        287      247     -40
## 6 UT      Ladakh                                  12        11      -1
## 7 UT      Lakshadweep                             2         1       -1
## 8 UT      Puducherry                             408      504      96
## 9 UT      Total (UTs)                           4315     4053    -262
## # ... with abbreviated variable names 1Suicide_2021, 2Variation
```

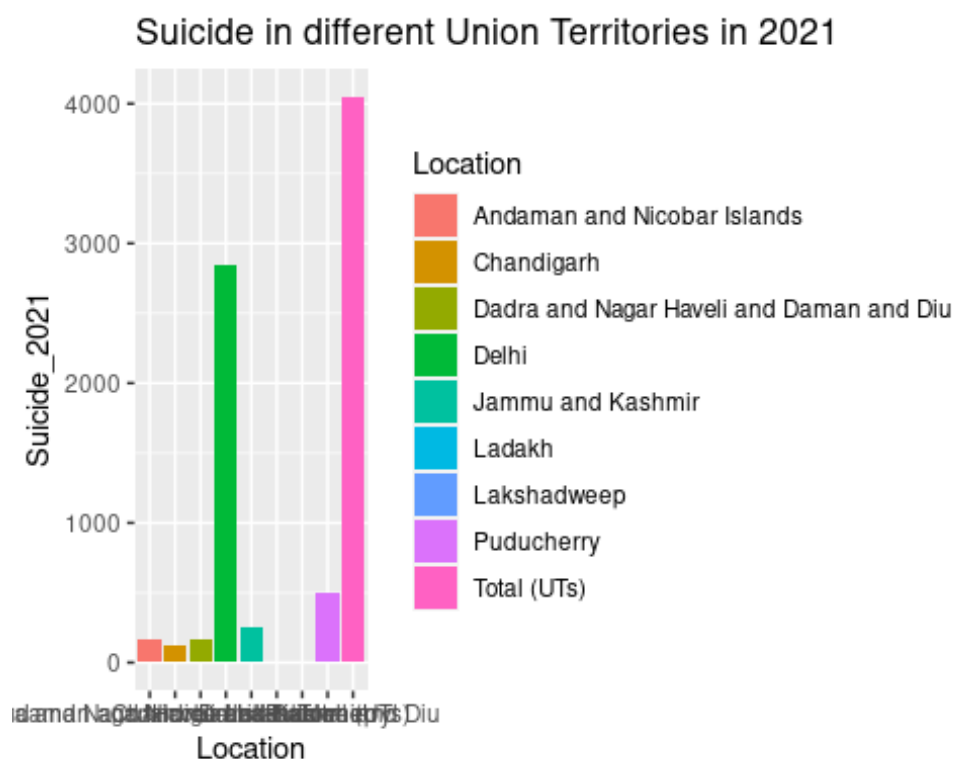
```
summary(ut)
```

```
##   Category      Location      Suicide_2020      Suicide_2021
## Length:9      Length:9      Min.   :   2.0      Min.   :   1.0
## Class :character Class :character 1st Qu.: 128.0      1st Qu.: 120.0
## Mode  :character Mode  :character Median : 180.0      Median : 171.0
##                                     Mean  : 958.9      Mean   : 900.7
##                                     3rd Qu.: 408.0      3rd Qu.: 504.0
##                                     Max.   :4315.0      Max.   :4053.0
##
## Variation
## Min.   : -302.00
## 1st Qu.: -40.00
## Median :  -8.00
## Mean   : -58.22
## 3rd Qu.:  -1.00
## Max.   :  96.00
```

```
ggplot(ut,aes(Location,Suicide_2020,fill=Location))+
  geom_bar(position="dodge",stat='identity')+
  labs(title = "Suicide in different Union Territories in 2020")
```

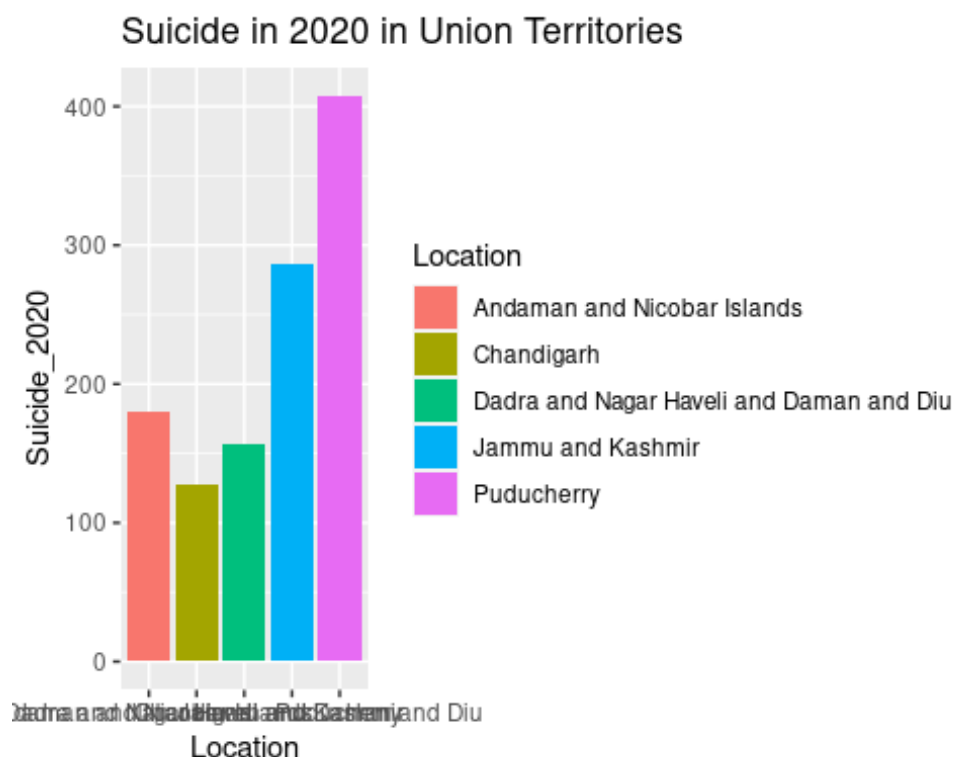


```
ggplot(ut,aes(Location,Suicide_2021,fill=Location))+
  geom_bar(position="dodge",stat='identity')+
  labs(title = "Suicide in different Union Territories in 2021")
```



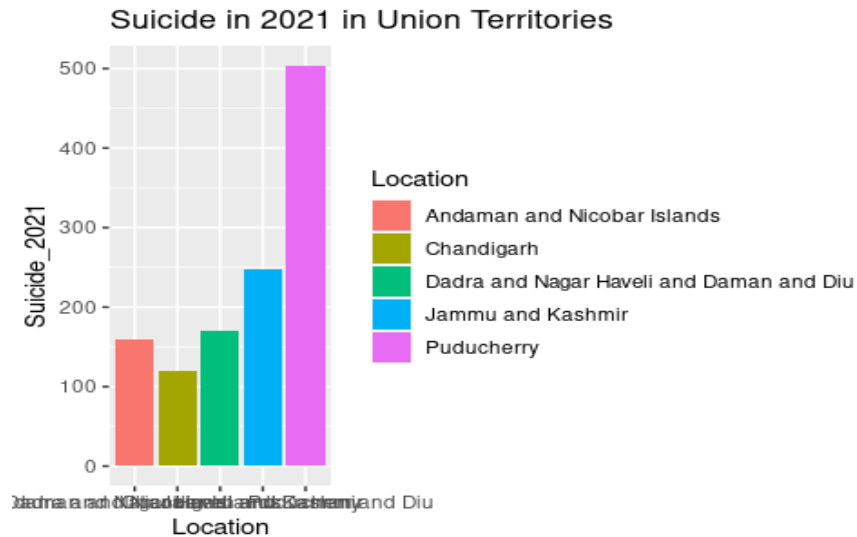
```
ut_2020 = filter(ut,Suicide_2020 >=128 & Suicide_2020<=408)
```

```
ggplot(ut_2020,aes(Location,Suicide_2020,fill=Location))+
  geom_bar(position="dodge",stat='identity')+
  labs(title = "Suicide in 2020 in Union Territories")
```



```
ut_2021 = filter(ut,Suicide_2021 >=120 & Suicide_2021<=504)
```

```
ggplot(ut_2021,aes(Location,Suicide_2021,fill=Location))+
  geom_bar(position="dodge",stat='identity')+
  labs(title = "Suicide in 2021 in Union Territories")
```



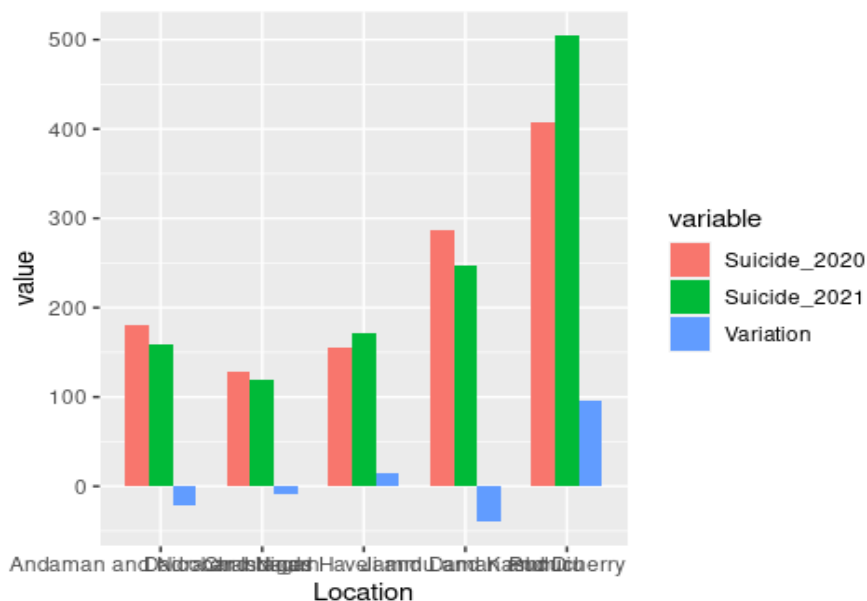
```
ut_1=inner_join(ut_2020,ut_2021)
```

```
## Joining with `by = join_by(Category, Location, Suicide_2020, Suicide_2021,`  
## Variation)`
```

```
ut_1 = ut_1[,-1]
```

```
ut_20_21 = melt(ut_1,id=c('Location'))
```

```
ggplot(ut_20_21)+
  geom_bar(aes(x=Location,y=value,fill=variable),
    stat="identity", position = "dodge", width = 0.7)
```



...