World suicide analysis

Eveline Xu xuehui

2020/12/1

Table of contents:

Data source

Suicide data. (2020). from https://www.who.int/teams/mental-health-and-substance-use/suicide-data

References

- World Health Organization. (2018). Suicide prevention，from http://www.who.int/mental_health/suicide-prevention/en/
- "Gender Differences In Suicide". 2020. En.Wikipedia.Org. https://en.wikipedia.org/wiki/Gender_differences_in_suicide#:~:text=Globally%2C%20death%20by%20suicide%20occurred,over%20the%20age%20of%2065.

Importing Data and Libraries

```
library(readr)
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(plotly)
```

```
Attaching package: 'plotly'
```

```
The following object is masked from 'package:ggplot2':

    last_plot
```

```
The following object is masked from 'package:stats':

    filter
```

```
The following object is masked from 'package:graphics':

    layout
```

```
suicideData <- read_csv("/Users/glameveline/RProjects/Fianl project/suicideData.csv")
```

```
Parsed with column specification:
cols(
  country = col_character(),
  year = col_double(),
  sex = col_character(),
  age = col_character(),
  suicides_no = col_double(),
  population = col_double()
)
```

```
head(suicideData)
```

country <chr>	year <dbl>	sex <chr>	age <chr>	suicides_no <dbl>	population <dbl>
Albania	1995	female	15-24 years	13	283500
Albania	1995	female	25-34 years	7	264000
Albania	1995	female	35-54 years	8	356400
Albania	1995	female	5-14 years	2	348700
Albania	1995	female	55-74 years	2	180400
Albania	1995	female	75+ years	2	40800

6 rows

Preprocessing

Dealling with missing values

```
summary(suicideData)
```

```
   country               year          sex                age
 Length:3492        Min.   :1995   Length:3492        Length:3492
 Class :character   1st Qu.:1995   Class :character   Class :character
 Mode  :character   Median :2005   Mode  :character   Mode  :character
                    Mean   :2004
                    3rd Qu.:2015
                    Max.   :2015

  suicides_no        population
 Min.   :    0.0   Min.   :     279
 1st Qu.:    1.0   1st Qu.:   71165
 Median :   14.0   Median :  387728
 Mean   :  205.4   Mean   : 1775099
 3rd Qu.:   95.0   3rd Qu.: 1371678
 Max.   :21706.0   Max.   :43509335
 NA's   :120       NA's   :396
```

Found missing values in suicides_no and population

```
table(is.na(suicideData$suicides_no))
```

```
FALSE  TRUE
 3372   120
```

```
table(is.na(suicideData$population))
```

```
FALSE  TRUE
 3096   396
```

Data analysis

- Rate of suicide per 100k in 1995, 2005 and 2015

```
suicide_year <- suicideData %>%
  filter(!is.na(population) & !is.na(suicides_no)) %>%
  select(year,suicides_no,population) %>%
  group_by(year) %>%
  summarise(suicide_ratio=round((sum(suicides_no)/sum(population))*100000,2))
```

```
`summarise()` ungrouping output (override with `.groups` argument)
```

```
suicide_year
```

year <dbl>	suicide_ratio <dbl>
1995	15.29
2005	12.94
2015	10.72

3 rows

```
ggplot(data=suicide_year,aes(x=year,y=suicide_ratio))+geom_line()+
  geom_point(colour = "red") +
  geom_text(aes(label = suicide_ratio, vjust =-2, hjust =0.5))+
  ylim(0,20)+
  labs(
    title = "World suicide rate trend over 20 years",
    subtitle = "1995,2005 and 2005",
    y = "Suicide Rate per 100K")
```

- suicide rate by age group from 1995 to 2015

```
suicide_age <-  suicideData %>%
  filter(!is.na(population) & !is.na(suicides_no)) %>%
  select(year,age,suicides_no,population) %>%
  group_by(age,year) %>%
  summarise(suicide_ratio=round((sum(suicides_no)/sum(population))*100000,2))
```

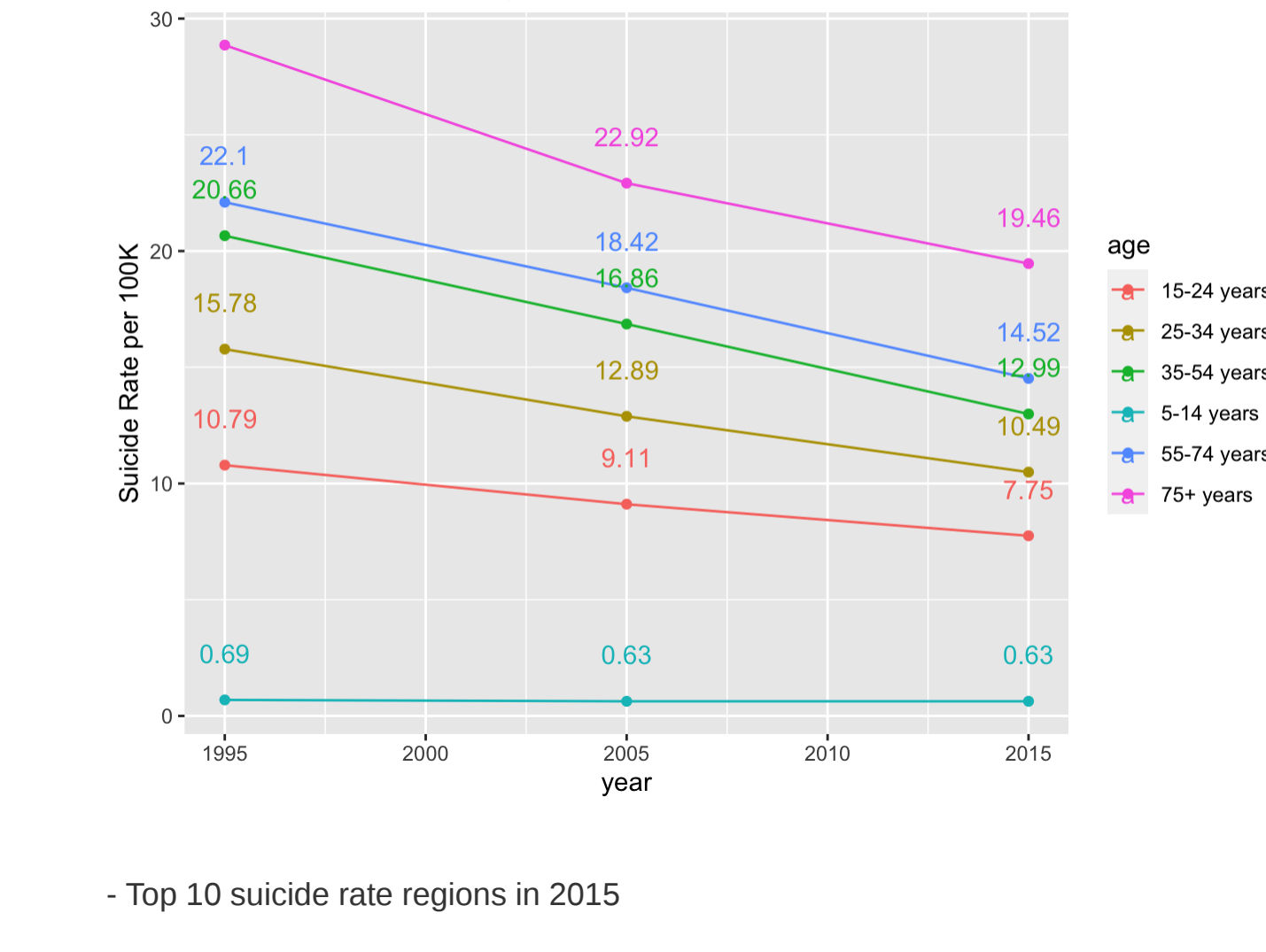
```
`summarise()` regrouping output by 'age' (override with `.groups` argument)
```

```
suicide_age
```

age <chr>	year <dbl>	suicide_ratio <dbl>
15-24 years	1995	10.79
15-24 years	2005	9.11
15-24 years	2015	7.75
25-34 years	1995	15.78
25-34 years	2005	12.89
25-34 years	2015	10.49
35-54 years	1995	20.66
35-54 years	2005	16.86
35-54 years	2015	12.99
5-14 years	1995	0.69

1-10 of 18 rows Previous 1 2 Next

```
ggplot(data=suicide_age,aes(x=year,y=suicide_ratio, group=age,color=age))+
  geom_line()+
  geom_point()+
        geom_text(aes(label = suicide_ratio, vjust =-2, hjust =0.5))+
  ggtitle("World suicide rate by age group in the previous 20 years")+
  labs(
    y="Suicide Rate per 100K"
  )
```



- Top 10 suicide rate regions in 2015

```
# top 10 per 100 k by region(country) in 2015
top10_suicide_country <- suicideData %>%
  filter(!is.na(population) & !is.na(suicides_no) & year=='2015') %>%
  select(country,suicides_no,population) %>%
  group_by(country) %>%
  summarise(suicide_ratio=round((sum(suicides_no)/sum(population))*100000,2)) %>%
  arrange(desc(suicide_ratio)) %>%
  head(10)
```

```
`summarise()` ungrouping output (override with `.groups` argument)
```

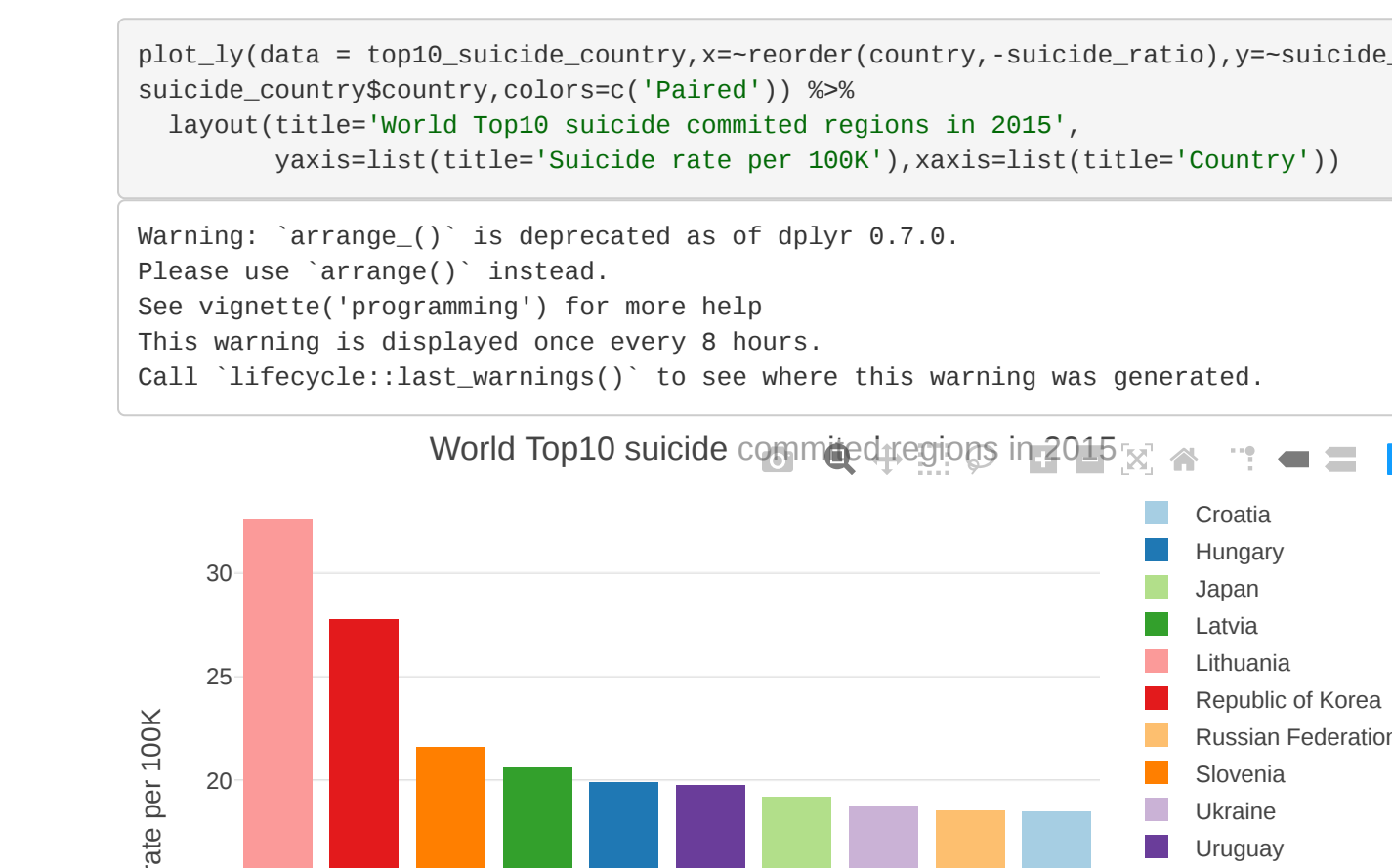
```
top10_suicide_country
```

country <chr>	suicide_ratio <dbl>
Lithuania	32.54
Republic of Korea	27.76
Slovenia	21.59
Latvia	20.63
Hungary	19.89
Uruguay	19.74
Japan	19.18
Ukraine	18.77
Russian Federation	18.56
Croatia	18.47

1-10 of 10 rows

```
plot_ly(data = top10_suicide_country,x=~reorder(country,-suicide_ratio),y=~suicide_ratio,type ='bar',color=top10_
suicide_country$country,colors=c('Paired')) %>%
  layout(title='World Top10 suicide commited regions in 2015',
         yaxis=list(title='Suicide rate per 100K'),xaxis=list(title='Country'))
```

```
Warning: `arrange_()` is deprecated as of dplyr 0.7.0.
Please use `arrange()` instead.
See vignette('programming') for more help
This warning is displayed once every 8 hours.
Call `lifecycle::last_warnings()` to see where this warning was generated.
```



- Choropleth map of suicide rate in the year of 2015

```
# Choropleth map of suicide rate in the year of 2015

suicide_country <- suicideData %>%
  filter(!is.na(population) & !is.na(suicides_no) & year=='2015') %>%
  select(country,suicides_no,population) %>%
  group_by(country) %>%
  summarise(suicide_ratio=round((sum(suicides_no)/sum(population))*100000,2))
```

```
`summarise()` ungrouping output (override with `.groups` argument)
```

```
suicide_country <- rename(suicide_country,region=country)

# importing world map
world_map <- map_data('world')
str(world_map)
```

```
'data.frame':   99338 obs. of  6 variables:
 $ long     : num  -69.9 -69.9 -69.9 -70 -70.1 ...
 $ lat      : num  12.5 12.4 12.4 12.5 12.5 ...
 $ group    : num  1 1 1 1 1 1 1 1 1 1 ...
 $ order    : int  1 2 3 4 5 6 7 8 9 10 ...
 $ region   : chr  "Aruba" "Aruba" "Aruba" "Aruba" ...
 $ subregion: chr  NA NA NA NA ...
```

```
# changing names of certain countries
world_map$region <- ifelse(world_map$region=='USA',"United States of America",world_map$region)
head(world_map,n=20)
```

	long <dbl>	lat <dbl>	group <dbl>	order <int>	region <chr>	subregion <chr>
1	-69.89912	12.45200	1	1	Aruba	NA
2	-69.89571	12.42300	1	2	Aruba	NA
3	-69.94219	12.43853	1	3	Aruba	NA
4	-70.00415	12.50049	1	4	Aruba	NA
5	-70.06612	12.54697	1	5	Aruba	NA
6	-70.05088	12.59707	1	6	Aruba	NA
7	-70.03511	12.61411	1	7	Aruba	NA
8	-69.97314	12.56763	1	8	Aruba	NA
9	-69.91181	12.48047	1	9	Aruba	NA
10	-69.89912	12.45200	1	10	Aruba	NA

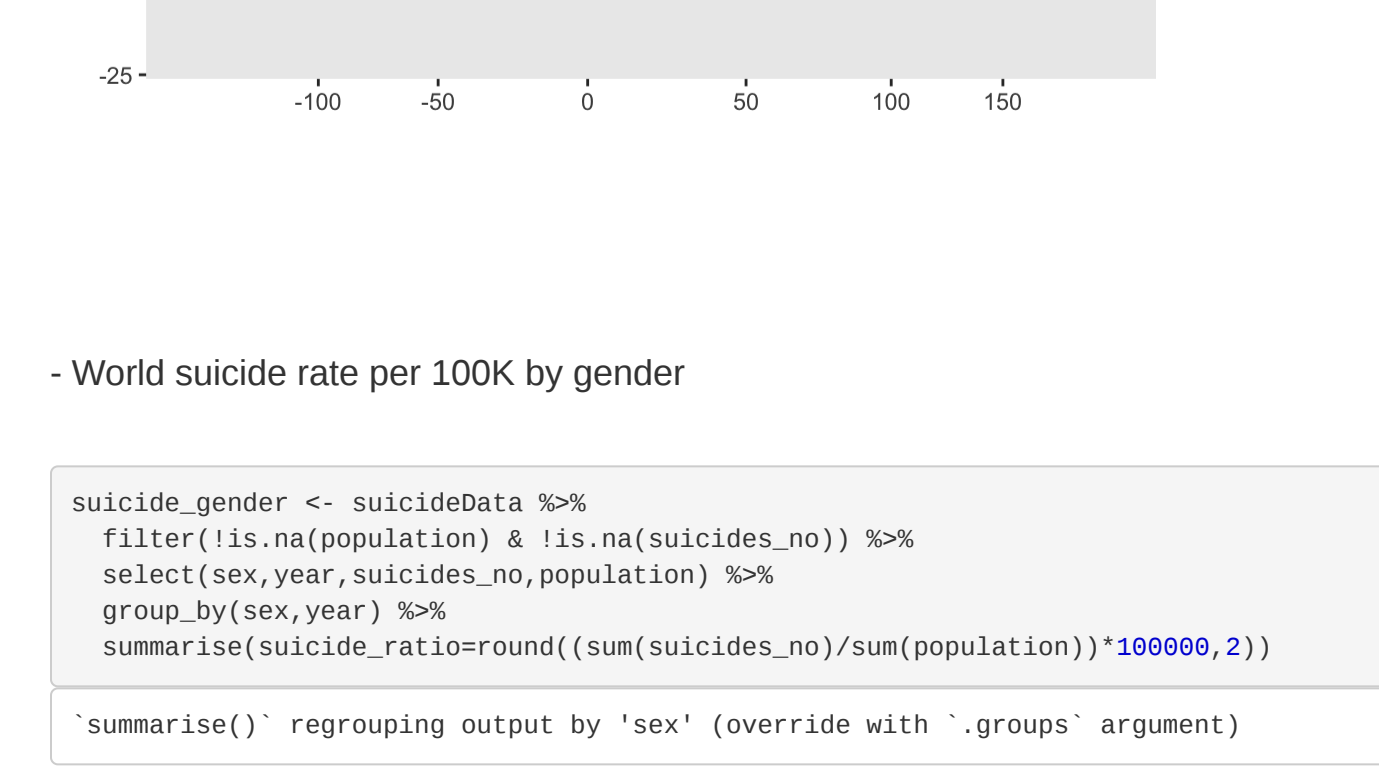
1-10 of 20 rows Previous 1 2 Next

```
# suicide map
world_suicide <- left_join(world_map,suicide_country,by='region')
head(world_suicide)
```

	long <dbl>	lat <dbl>	group <dbl>	order <int>	region <chr>	subregion <chr>	suicide_ratio <dbl>
1	-69.89912	12.45200	1	1	Aruba	NA	9.12
2	-69.89571	12.42300	1	2	Aruba	NA	9.12
3	-69.94219	12.43853	1	3	Aruba	NA	9.12
4	-70.00415	12.50049	1	4	Aruba	NA	9.12
5	-70.06612	12.54697	1	5	Aruba	NA	9.12
6	-70.05088	12.59707	1	6	Aruba	NA	9.12

6 rows

```
# vidualization
arrange(world_suicide, order) %>%
  ggplot(aes(long, lat, group = group, fill =suicide_ratio)) +
  geom_polygon() +
  xlim(-110,150)+ylim(-20,90)+
  coord_map("polyconic") +
  theme(panel.grid =element_blank())+
  scale_fill_continuous(low = "green", high = "red")+
  labs(
    y='',
    x=''
  )
```



- World suicide rate per 100K by gender

```
suicide_gender <- suicideData %>%
  filter(!is.na(population) & !is.na(suicides_no)) %>%
  select(sex,year,suicides_no,population) %>%
  group_by(sex,year) %>%
  summarise(suicide_ratio=round((sum(suicides_no)/sum(population))*100000,2))
```

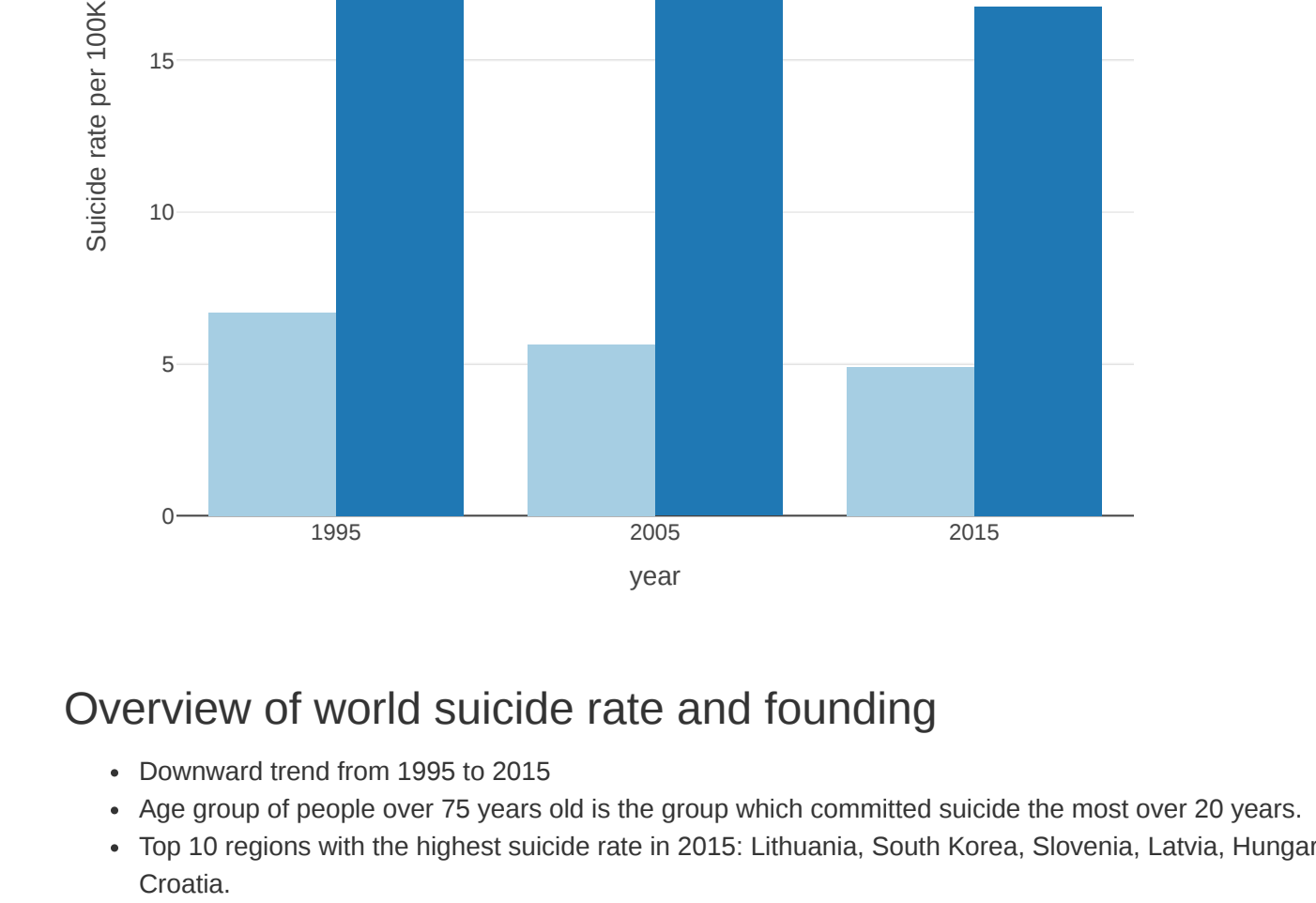
```
`summarise()` regrouping output by 'sex' (override with `.groups` argument)
```

```
suicide_gender
```

sex <chr>	year <dbl>	suicide_ratio <dbl>
female	1995	6.68
female	2005	5.63
female	2015	4.90
male	1995	24.38
male	2005	20.59
male	2015	16.78

6 rows

```
# vidualization
plot_ly(data = suicide_gender,x=~year,y=~suicide_ratio,type ='bar',color=suicide_gender$sex,colors=c('Paired')) %
>%
  layout(title='World suicide rate per 100K by gender',
         yaxis=list(title='Suicide rate per 100K'))
```



Overview of world suicide rate and founding

- Downward trend from 1995 to 2015
- Age group of people over 75 years old is the group which committed suicide the most over 20 years.
- Top 10 regions with the highest suicide rate in 2015: Lithuania, South Korea, Slovenia, Latvia, Hungary, Uruguay, Japan, Ukraine, Russia, Croatia.
- More male die from suicide than women.

Interpretation and reasonings

Why more male die from suicide than women?

*This is also known as the gender paradox in suicide.

- Suicide attempts are between two and four times more frequent among females.
- Researchers have partly attributed the difference between attempted and completed suicides among the sexes to males using more lethal means to end their lives.

Why most of countries in top10 suicide rate are in East Europe?

- Economic crisis after collapse of USSR.
- Civil war.

Suicide prevention

- September 10 has been observed as 'World Suicide Prevention Day' in partnership with the International Association for Suicide Prevention and the World Health Organization.
- South Korea Suicide Hotlines http://www.suicide.org/hotlines/international/south-korea-suicide-hotlines.html
- Talk therapy
- Suicide Survivors are Heroes