US Gun Murders-2010

Harkanwal Kohli

GitHub: 'https://github.com/HarkanwalSingh761315'

Twitter: 'https://twitter.com/harkanwalkohli'

```
library(dslabs) ##importing dslabs library for datasets
library(tidyverse) ##importing library tidyverse for data manipulation
```

```
## Warning: package 'tidyverse' was built under R version 4.0.3
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## Warning: package 'ggplot2' was built under R version 4.0.3
```

```
## Warning: package 'readr' was built under R version 4.0.3
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)##importing library ggplot2 to plot the graphs
library(ggrepel)##importing the library ggrepel to avoid point and text overlap
```

```
## Warning: package 'ggrepel' was built under R version 4.0.3
```

```
data(murders) ##Loading data murders
df <- murders ## adding dataset to df
```

```
names(df) ##TO know the column names
```

```
## [1] "state"      "abb"        "region"     "population" "total"
```

```
str(df) ##To know the structure of data frame (SUMMARY FUNCTION)
```

```
## 'data.frame':    51 obs. of  5 variables:
##  $ state     : chr  "Alabama" "Alaska" "Arizona" "Arkansas" ...
##  $ abb       : chr  "AL" "AK" "AZ" "AR" ...
##  $ region    : Factor w/ 4 levels "Northeast","South",..: 2 4 4 2 4 4 1 2 2 2 ...
##  $ population: num  4779736 710231 6392017 2915918 37253956 ...
##  $ total     : num  135 19 232 93 1257 ...
```

```
summary(df) #To know the summary of df (SUMMARY FUNCTION)
```

```
##     state               abb                 region     population
##  Length:51          Length:51          Northeast    : 9   Min.   :  563626
##  Class :character   Class :character   South        :17   1st Qu.: 1696962
##  Mode  :character   Mode  :character   North Central:12   Median : 4339367
##                                        West         :13   Mean   : 6075769
##                                                           3rd Qu.: 6636084
##                                                           Max.   :37253956
##      total
##  Min.   :   2.0
##  1st Qu.:  24.5
##  Median :  97.0
##  Mean   : 184.4
##  3rd Qu.: 268.0
##  Max.   :1257.0
```

```
head(df,10) ##to see the top 10 entries
```

```
##                  state abb    region population total
## 1              Alabama  AL     South    4779736   135
## 2               Alaska  AK      West     710231    19
## 3              Arizona  AZ      West    6392017   232
## 4             Arkansas  AR     South    2915918    93
## 5           California  CA      West   37253956  1257
## 6             Colorado  CO      West    5029196    65
## 7          Connecticut  CT Northeast    3574097    97
## 8             Delaware  DE     South     897934    38
## 9  District of Columbia  DC     South     601723    99
## 10              Florida  FL     South   19687653   669
```

```
df_mutate <- df %>%
  mutate(murderRate = total/population*10^5) ##adding new column murder rate

USA_murder_rate <- df_mutate %>%
  summarise(USA_rate = sum(total)/sum(population)*10^6) %>%
  pull(USA_rate)
```

```
MurdersAndPopulation_plot <- df_mutate %>%
  ggplot(aes(x=population/10^6,y=total,na.rm=TRUE)) ##defining aesthetic variables to ggplot

MurdersAndPopulation_plot <- MurdersAndPopulation_plot+ ##defining the type of graph needed
  geom_point(aes(color=region))

MurdersAndPopulation_plot <- MurdersAndPopulation_plot + ##defining the axis scales
  scale_x_continuous(trans = "log10") +
  scale_y_log10()

MurdersAndPopulation_plot <- MurdersAndPopulation_plot+ ##Defining the lables
  labs(x="Population per Million",
       y= "Murders",
       title="Average Gun Murders",
       subtitle="USA",
       caption="Year-2010")

MurdersAndPopulation_plot <- MurdersAndPopulation_plot+ ##plotting the average rate line
  geom_abline(intercept = log10(USA_murder_rate),lty=3,color="grey")

MurdersAndPopulation_plot <- MurdersAndPopulation_plot+## to avoid point and text overlap
  geom_text_repel(aes(label=abb, color=region))
```

```
MurdersAndPopulation_plot ##viewing the graph
```