Quiz 2

Zeshi Feng

2025-09-16

Data: In a Washington Post article from October 6, 2015, titled "Zero correlation between state homicide rate and state gun laws", Eugene Volokh examined the relationship between the state total homicide rate and state gun laws.

To do that, Volokh used the '**Brady score**', which represents how difficult it is to obtain a gun in a state. A low Brady score means a low level of gun restrictions (so it is easier to obtain a gun), and a high score means that it is harder to get a gun (i.e. there are stricter gun laws).

The data set, available in the file `gun.csv`, contains the following variables:

- Jurisdiction: State/jurisdiction name
- Homicide rate: Number of **intentional** homicide death per 100,000 people
- Gun accident rate: Number of gun accident death per 100,000 people
- Brady score: The Brady score represents how difficult it is to obtain a gun in a state.
- Brady grade: Another interpretation of Brady Score

Answer the following questions:

a. Load the data `gun.csv` into R and store it as **gunData**. Check the dimensions and structure of the dataset. Identify the data type of each variables.

```
gunData <- read.csv("gun.csv")
dim(gunData)
```

```
## [1] 51  5
```

```
str(gunData)
```

```
## 'data.frame':    51 obs. of  5 variables:
##  $ Jurisdiction     : chr  "Alabama" "Alaska" "Arizona" "Arkansas" ...
##  $ Homicide.rate    : num  7.1 4.1 5.5 5.9 5 3.1 4.1 13.9 6.2 5.2 ...
##  $ Gun.accident.rate: num  0.44 0 0.1 0.41 0.08 0.18 0 0 0 0.13 ...
##  $ Brady.score      : num  3.5 -7 -8 1 75 14.5 70 50 34.5 3 ...
##  $ Brady.grade      : chr  "D-" "F" "F" "F" ...
```

"The dataset has 51 rows and 5 columns, the data type are showing above"

b. Add a logical variable **Homicide_index** to the data frame, takes the value TRUE if Homicide.rate is greater than 4.5, and FALSE otherwise. Verify the class of this variable is logical. Use the `table()` function to show the number of TRUE's and the number of FALSE's.

```
gunData$Homicide_index <- gunData$Homicide.rate > 4.5
class(gunData$Homicide_index)
```

```
## [1] "logical"
```

```
table(gunData$Homicide_index)
```

```
##
## FALSE  TRUE
##    30    21
```

c. Add a new categorical variable **bScoreCategory** to the data frame, which divides the Brady scores into two categories: high (H) if the Brady score is greater than 10, and low (L) if it is less than or equal to 10. Make sure the variable is a factor with two levels, "H" and "L". Display the summary statistics of the new variable.

```
gunData$bScoreCategory <- ifelse(gunData$Brady.score > 10, "H", "L")
gunData$bScoreCategory <- factor(gunData$bScoreCategory, levels = c("H", "L"))
class(gunData$bScoreCategory)
```

```
## [1] "factor"
```

d. Add a new variable named **Sum** to the data frame gunData by calculating the sum of `Homicide.rate` and `Gun.accident.rate`. Note that this variable represents the **total homicide rate**, i.e., total intentional homicide plus accidental gun death rates.

```
gunData$Sum <- gunData$Homicide.rate + gunData$Gun.accident.rate
head(gunData[, c("Jurisdiction", "Homicide.rate", "Gun.accident.rate", "Sum")])
```

```
##   Jurisdiction Homicide.rate Gun.accident.rate  Sum
## 1      Alabama           7.1              0.44 7.54
## 2       Alaska           4.1              0.00 4.10
## 3      Arizona           5.5              0.10 5.60
## 4     Arkansas           5.9              0.41 6.31
## 5   California           5.0              0.08 5.08
## 6     Colorado           3.1              0.18 3.28
```

e. Identify the top 3 states with the highest total homicide rates and retrieve their corresponding Brady score, Homicide.rate, and Gun.accident.rate. (Hint: You may sort the total homicide rates for all states in descending order, and use the third-highest total homicide rate as a cutoff to extract the relevant rows.)

```
sortedData <- gunData[order(-gunData$Sum), ]
cutoff <- sortedData$Sum[3]
```

```
topStates <- subset(gunData, Sum >= cutoff)
topStates
```

```
##    Jurisdiction Homicide.rate Gun.accident.rate Brady.score Brady.grade
## 8          D.C.          13.9              0.00          50           B
## 19    Louisiana          10.8              0.75          -2           F
## 25  Mississippi           7.4              0.65          -4           F
##    Homicide_index bScoreCategory   Sum
## 8            TRUE              H 13.90
## 19           TRUE              L 11.55
## 25           TRUE              L  8.05
```

f. Find all the states/jurisdictions which have weak restrictions for buying guns (Brady score < 0), and also have a low total homicide rate (< 4). Display the corresponding state/jurisdiction name, Brady score, and total homicide rate.

```
weak_low <- subset(gunData, Brady.score < 0 & Sum < 4)
weak_low[, c("Jurisdiction", "Brady.score", "Sum")]
```

```
##    Jurisdiction Brady.score  Sum
## 17       Kansas        -4.0 3.10
## 27      Montana        -3.0 3.13
## 42 South Dakota        -4.5 3.00
## 45         Utah        -2.0 1.80
## 46      Vermont        -4.0 1.30
## 51      Wyoming        -5.0 2.40
```