Week-10:diary entry

Wang Renhe

25-10-2023

Question 1: What is the topic that you have finalized?

Solution: current topic: Homicide rate in US from 2001-2010. It can help us understand of and prevent future crime, for the welfare of the society.

Question 2. What are the data sources that you have curated so far?

Solution: I get this data set from Kaggle

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
homicide <- read.csv("database1.csv")
head(homicide)
```

```
##   Record.ID Agency.Code Agency.Name      Agency.Type      City  State Year
## 1         1     AK00101   Anchorage Municipal Police Anchorage Alaska 1980
## 2         2     AK00101   Anchorage Municipal Police Anchorage Alaska 1980
## 3         3     AK00101   Anchorage Municipal Police Anchorage Alaska 1980
## 4         4     AK00101   Anchorage Municipal Police Anchorage Alaska 1980
## 5         5     AK00101   Anchorage Municipal Police Anchorage Alaska 1980
## 6         6     AK00101   Anchorage Municipal Police Anchorage Alaska 1980
##     Month Incident             Crime.Type Crime.Solved Victim.Sex Victim.Age
## 1 January        1 Murder or Manslaughter          Yes       Male         14
## 2   March        1 Murder or Manslaughter          Yes       Male         43
## 3   March        2 Murder or Manslaughter           No     Female         30
## 4   April        1 Murder or Manslaughter          Yes       Male         43
## 5   April        2 Murder or Manslaughter           No     Female         30
## 6     May        1 Murder or Manslaughter          Yes       Male         30
##                      Victim.Race Victim.Ethnicity Perpetrator.Sex
## 1 Native American/Alaska Native          Unknown            Male
## 2                          White          Unknown            Male
```

```
## 3 Native American/Alaska Native          Unknown         Unknown
## 4                          White          Unknown            Male
## 5 Native American/Alaska Native          Unknown         Unknown
## 6                          White          Unknown            Male
##   Perpetrator.Age              Perpetrator.Race Perpetrator.Ethnicity
## 1              15 Native American/Alaska Native                Unknown
## 2              42                          White                Unknown
## 3               0                        Unknown                Unknown
## 4              42                          White                Unknown
## 5               0                        Unknown                Unknown
## 6              36                          White                Unknown
##   Relationship        Weapon Victim.Count Perpetrator.Count Record.Source
## 1 Acquaintance  Blunt Object            0                 0           FBI
## 2 Acquaintance Strangulation            0                 0           FBI
## 3      Unknown       Unknown            0                 0           FBI
## 4 Acquaintance Strangulation            0                 0           FBI
## 5      Unknown       Unknown            0                 1           FBI
## 6 Acquaintance         Rifle            0                 0           FBI
```

Week10-Continue from Week9

What is the question that you are going to answer?

How does the victim-murderer relationship, and the age of perpetrators affect the choice of weapons used during homicides?

Why is this an important question?

A research conducted by American Journal of Preventive Medicine suggest a rising trend in intimate partner homicides, as well as a higher occurrence of homicides involving friends, acquaintances, and even strangers. Those findings underscore the importance of relevant laws.

From Bailey et al.(2023), urban areas exhibit a higher propensity for the escalation of gun violence as compared to other weapon-related incidents, identify the pivotal age group can effectively establish deployment strategies, so can prevent future trauma.

According to United Nations, reduce the crime rate will reduce potential harmful effects on individuals and society.

Which rows and columns of the dataset will be used to answer this question?

Incident, Perpetrator Age, Relationship, Weapon

Include the challenges and errors that you faced and how you overcame them

First, data entry errors may occur when there are missing values or incorrect data entries, which can distort the dataset and affect the accuracy of the results. Hence, it is necessary to address missing values by removing them from the dataset before conducting any analysis. Secondly, if outliers exist, the unusual or erroneous data points can significantly impact statistical analysis. Therefore, I will try to highlight those values and see if I should exclude or include them in the analysis.Moreover, incorrect data transformations can lead to inaccurate results because linear regression can be used with a combination of categorical and numerical values. To address this, it's essential to apply suitable data transformation techniques to incorporate categorical variables correctly into the analysis.