Week-12:Diary entry

Wang Renhe

09-11-2023

Week9

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

Solution: current topic: Homicide rate in US in 2014. 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 do the victim-murderer relationship and the age, and the sex of perpetrators impact the choice of weapons used in 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?

Perpetrator Age, Relationship, Weapon, Perpetrator Sex

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.

Week11

(1) List the visualizations that you are going to use in your project (Answer: What are the variables that you are going to plot? How will it answer your larger question?)

ggplot (bar charts/point charts) of two sets of variables:

1. "Relationship" and "Weapon": I will plot separate graphs for weapon types and the different murder-victim relationships. After plotting graphs for all weapons, I will combine all and come up with overall bar chart using ggplot

By knowing this, this can show me if there is any pattern between the victim-murder relationship and weapon use during homicide. And this will answer my larger question. If yes, then we will know how to comprehense the law and restrict consumption for certain weapons. Meanwhile, we will also being informed what kind of weapon can be avoid, i.e. knife should be kept away or only accessible by them, so that they can protect themselves.If no then, it means no significance relationship between, hence, all weapons should be treated equally dangerous

2. "Perpetrator Age" and "Weapon". I will plot bar graph to see if there is weapon preference about different age ranges during homicide. It will answer the larger question. If yes, then maybe younger generation prefer one weapon over the other.

I will create a density plot to determine which age group has the highest occurrence, which may help explain which weapons' usage will surge.

3. "sex" of perpetrator and "weapon" used.

See if there is certain preference towards weapons use for different sex/gender. I will plot separate graph for different sex(male/female), with "sex" in x-axis and "weapon" type in y-axis.Then provide overall bar chart with both sexes. It will answer my larger question to see which weapon is commonly used for males/females. For instance, females may prefer knives more, males may prefer guns more etc.

(2) How do you plan to make it interactive? (Answer: features of ggplot2/shiny/markdown do you plan to use to make the story interactive)

Utilizing ggplot2 for the purpose of viewing various factors such as gender, age groups, relationship statuses, and their association with different types of weapon usage.

(3) What concepts incorporated in your project were taught in the course and which ones were self-learnt? (Answer: Create a table with topics in one column and Weeks in the other to indicate which concept taught in which week is being used. Leave the entry of the Week column empty for self-learnt concepts)

```r
df <- data.frame(
  Topic = c ("insert picture", "change variable type","create a new list","Choosing rows or columns","a
  Week = c ("Week1","Week3","Week3","Week4","Week4","Week4","Week4","Week2/7","Week7","Week7","Week8","
)

df
```

```
##                               Topic         Week
## 1                    insert picture        Week1
## 2               change variable type        Week3
## 3                  create a new list        Week3
## 4           Choosing rows or columns        Week4
## 5                     arrange columns        Week4
## 6  Combining two or more operations        Week4
## 7                         filter data        Week4
## 8                             ggplot2      Week2/7
## 9                           histogram        Week7
## 10                            mapping        Week7
## 11                   User Interface        Week8
## 12               Interactive ggplot OnlineSource
```

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

First, to eliminate the invalid or blank.

Secondly, the data set may be too large and too complex, therefore I need to break down the complex dataset into smaller subsets or focus on variables of interest. Besides, use the correct way to run the code.

Thirdly, given that my data covers a wide time range, it's important to acknowledge that data volume and accuracy may not have been as reliable several decades ago when technology was less advanced. Nevertheless, I will make every effort to ensure that I select a comparable amount of data for analysis in both 2014 and 2004, to get relatively fair comparison