Week-13:Diary entry

Wang Renhe

11-11-2023

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("dataset3.csv")
head(homicide)
```

```
##   Year Perpetrator.Age Relationship       Weapon Perpetrator.Sex
## 1 2014              19     Neighbor      Firearm            Male
## 2 2014              24 Acquaintance      Firearm            Male
## 3 2014              39     Stranger Blunt Object            Male
## 4 2014              22   Girlfriend      Firearm            Male
## 5 2014              38       Family        Knife            Male
## 6 2014              26 Acquaintance      Firearm            Male
```

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?

Year, 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. Analyzing the relationship between "Relationship" and "Weapon" in separate graphs for different weapon types and murder-victim relationships can offer insights into potential patterns in homicide incidents.

By creating individual graphs for each weapon and then consolidating the data into an overall bar chart using ggplot, you can discern whether there is a significant correlation between the victim-murderer relationship and the choice of weapon.

This analysis can help address the broader question of whether certain weapons should be subject to legal restrictions or access control. For example, if the data reveals a strong connection between, say, knives and a specific victim-murderer relationship, it might suggest the need for stricter regulation on knife access. Conversely, if there is no statistically significant relationship, it may indicate that all weapons should be treated equally in terms of their potential danger.

2. Investigating the relationship between "Perpetrator Age" and the choice of weapon through bar graphs can shed light on potential age-related weapon preferences in homicide cases.

By examining the data, you can determine whether there's a tendency for certain age groups to favor specific weapons over others. Additionally, creating a density plot can help identify the age group with the highest frequency of involvement in homicide incidents. This information can contribute to understanding which age groups might be more inclined to use particular weapons, providing valuable insights into potential motivations or patterns in weapon usage.

3. Exploring the connection between the "sex" of the perpetrator and the type of "weapon" used through separate graphs for male and female perpetrators, followed by an overall bar chart, can reveal whether there are gender-based preferences in weapon selection during homicides.

By analyzing the data in this way, you can ascertain which weapons are more commonly associated with males and females, potentially uncovering gender-specific trends in weapon usage. For example, if the analysis indicates that females tend to prefer knives while males favor guns, this information could inform strategies for prevention or intervention tailored to specific gender groups.

(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)

```
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
## 13                        Casewhen    NewSource
```

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

Week13

(1) What is the theme of your data story?

theme: Understanding the Complex Interplay of Factors in Homicides. research question: How do the victim-murderer relationship and the age, and the sex of perpetrators impact the choice of weapons used in homicides?

(2) Why is it important to address this question?

Exploring the intricate relationships between victim-murderer dynamics, perpetrator age, sex(exclude intersex), and weapon choice can provide crucial insights into the contributing factors of homicides. This understanding is pivotal for designing targeted crime prevention strategies that address specific needs and risk factors associated with different scenarios, potentially aiding in the prevention of future crimes (Hsu et al., 2022)

By identifying these preferences, it emphasizes the significance of data-driven decision-making in shaping policies related to weapon control and access (Welsh & Farrington, 2012). This highlights the necessity for nuanced regulations that account for the multifaceted nature of homicides, moving away from implementing uniform measures.

Understanding the complex interplay of factors in homicides can enhance public awareness regarding the range of circumstances and motivations behind such events, fostering more informed dialogues and challenging misconceptions (Lewis, 2017).

Tailoring interventions to specific age groups and sex groups (exclude intersex) is crucial. Identifying significant weapon preferences in homicides allows law enforcement to focus on particular types of murder, whether within specific age or sex groups. This targeted focus enables specialized attention and consideration to address these distinct concerns effectively.

(3) Why do you think the data sources that you have curated can help you answer the question?

I have focused on narrowing down the dataset for a more targeted investigation. Instead of considering the entire range from 1980 to 2014, I have opted to concentrate solely on the most recent data from the year 2014. Additionally, to address my specific research question about the relationship between perpetrator age, relationship to the victim, and the weapon used in homicides, I have filtered out irrelevant variables like race and month.

With a streamlined dataset that includes the key columns of "Perpetrator Age," "Relationship," "Weapon," and "Perpetrator Sex," I can conduct pairwise comparisons between these variables. By creating graphs that intersect two variables at a time, I can visualize the relationships between them. This approach allows me to gain a comprehensive view of the proportions for each variable and how they influence the choice of

weapons in homicide cases. Ultimately, these insights will be instrumental in addressing the broader research question I have set out to answer.

(4) What are the insights from the data and how are they depicted in plots?

To answer: How do the victim-murderer relationship and the age, and the sex of perpetrators impact the choice of weapons used in homicides?

While there is no significant difference in the use of weapons across different victim-murder relations and sexes, the top three weapons remain consistent. The proportions of each weapon choice reveal noteworthy findings. Strangers, more than half the time, opt for handguns due to their accessibility, while closer relationships tend to involve strangulation, a more intimate method carried out by hand or rope or plastic bags. As relationships become closer, the preference for handguns slightly diminishes, giving way to more intimate or accessible weapons. Amid family dynamics, there is not a discernible preference for a specific weapon; instead, blunt objects, knives, and other firearms are frequently utilized. Despite the US Congress allocating $25 million for gun violence prevention (Rosenberg, 2020), it's imperative to acknowledge that focusing solely on funding and legislative measures in this domain might not be the most comprehensive strategy for preventing murder cases. Exploring other potential aspects related to weapon use, such as gender, age, and neighborhood safety, as suggested by Houtsma & Raines (2023), is vital to ensure a safer society in the future. Regarding sex variations, females exhibit a higher tendency to use knives compared to males. With respect to perpetrator age impacting weapon choice, handguns and blunt objects show no age variation, being utilized across all age groups. Rifles, suffocation, and strangulation tend to occur more frequently within the middle and earlier age groups. This is a way to further reduce murder cases.

(5) How did you implement this entire project? Were there any new concepts that you learnt to implement some aspects of it?

I've generated three graphs by employing different sets of variables, drawing on concepts from my existing knowledge base. Through filtering, selection, and recoding of specific variables, I've crafted plots using ggplot. However, to augment the interactivity of these plots, I've explored a new concept known as ggplotly.