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.

Data Story

Larger question

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

Why is it important to address this question?

1. 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)

2. 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.

3. 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).

4. 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.

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.

After clearing up the data, including filter the year to only 2014 and a remove all missing values or "Not Clear" values, new dataset is formed.

```
##   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
```

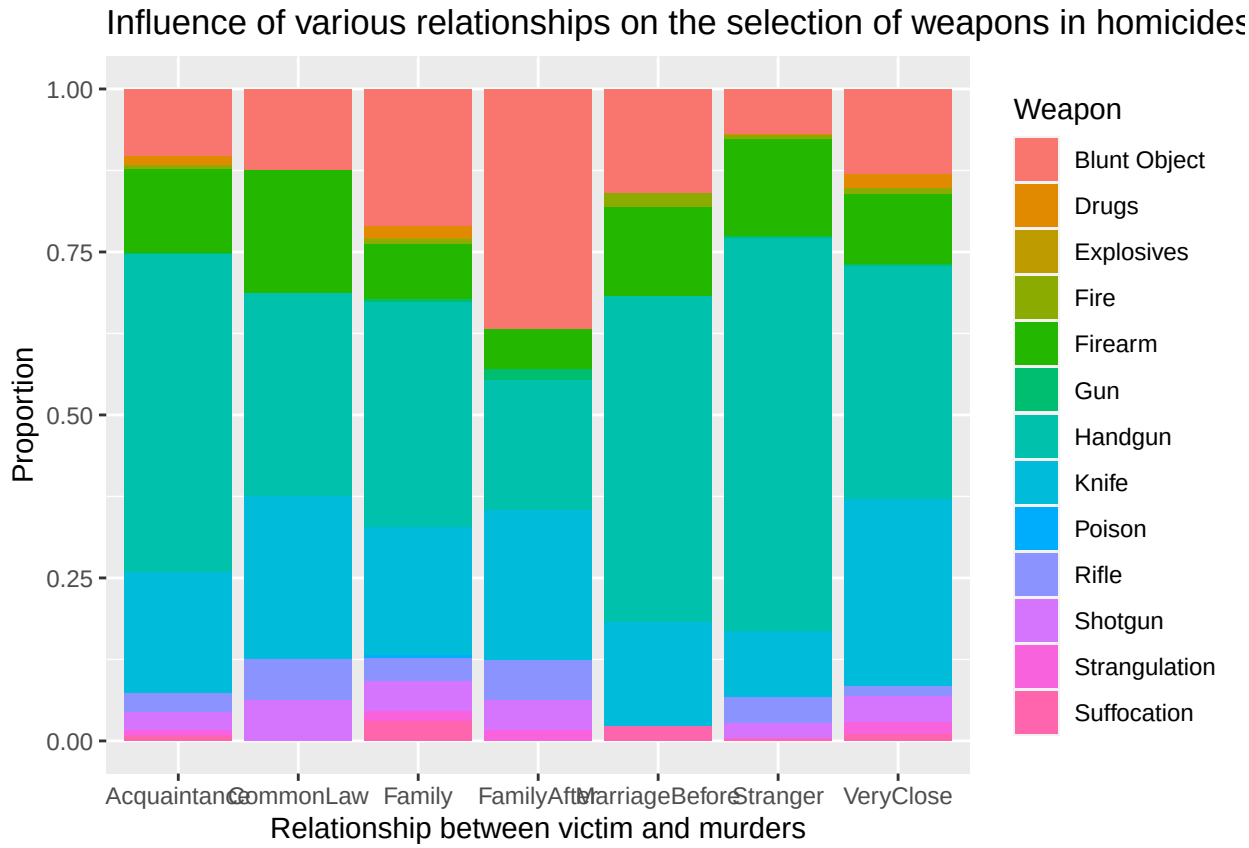
This table exclusively presents the variables I have specifically selected for assessing their influence on weapon choice in homicide cases.

To simplify the comparison process, I employed the "case when" to categorize the varied relationships in the "Relationship" variable into clear and distinct groups. The first category, labeled "Family," encompasses close blood relationships such as "father", "mother", "daughter", "son", "wife", "brother", "sister", and "husband". The classification of "common-law husband" and "common-law wife" acknowledges their distinct dynamics, characterized by a more open and less conventional relationship than formal marriages (Grossbard & Vernon, 2014). Furthermore, Members of reconstituted families, such as "Stepdaughter," "Stepson," and "Stepfather," are now collectively classified under the category "FamilyAfter." This grouping acknowledges that their sentiments and dynamics within the family may differ from those in the original family structure. Moreover, individuals with close but non-blood ties, where significant responsibilities may not be present, such as friends and boyfriends/girlfriends, are now categorized within the "VeryClose" group. Nevertheless, "Ex-wife" and "Ex-husband" represent former relationships that no longer carry legal responsibilities. "Acquaintance" combines "Employee," "Employer," and "Neighbor," acknowledging their transient or contractual nature. Lastly, "Stranger" designates individuals with whom there is no prior connection. This recoding provides a clearer framework for comparing the impact of different relationship types on the variable "Weapon" in our analysis.

```
##   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
##   RelationshipNew
## 1    Acquaintance
## 2    Acquaintance
## 3        Stranger
## 4       VeryClose
## 5          Family
## 6    Acquaintance
```

Relationship and its impact to weapon choice

In order to use the latest version of variable "Relationship", I add a new column named "RelationshipNew" using "mutate" function.



The bar chart above illustrates the varying proportions of different weapons used in homicides across different relationship categories. In the "Acquaintance" category, approximately 49% of homicides involve handguns, followed by 19% using knives, and 13.5% involving firearms. In the context of common-law relationships, characterized by a more open marriage structure, 25% of cases involve handguns, 25% knives, and 18.8% firearms. Although shotguns and rifles have higher percentages than in the "Acquaintance" category, their overall occurrence is relatively low at 6.25% each. Family-related homicides show a distribution of 28.1%

involving handguns, 22.5% blunt objects, and 19% with victims being killed by knives. For reconstituted families, 40.5% of homicides use blunt objects as the main weapon, followed by approximately 22.5% using knives and 22% using handguns. In cases involving ex-wives and ex-husbands, 44% of homicides involve handguns, with an equal proportion (18.8%) for knives, blunt objects, and firearms. Homicides involving strangers exhibit a notable 64% using handguns, 19% using firearms, and 12% using knives. Within the "VeryClose" relationship category, the top three weapons used are 37.5% handguns, 27% knives, and 14% blunt objects. It is worth noticing that "firearm" is a generic term and not a specific type of weapon.

In summary, while the choice of the most frequently used weapons remains consistent across all homicide cases, the proportions vary significantly depending on the nature of the relationship. In cases involving strangers and less familiar relationships, handguns play a prominent role, likely due to their accessibility and ease of concealment. In California, obtaining a gun involves obtaining a hunter education certificate and undergoing a background check, a process that can be completed relatively swiftly (Wertz et al., 2018). Conversely, as relationships grow closer, as seen in family-related homicides, the distribution of weapon choices becomes more evenly spread. This suggests that in closely related individuals, there is less favoritism towards specific weapons, necessitating equal consideration for each type. Notably, family homicides showcase a distinct prevalence of blunt object use, indicating a unique pattern in weapon selection within familial contexts (Mervosh, 2019). Furthermore, the increased occurrence of strangulation in closer relationships can be attributed to the intimate nature of this method, underscoring the proximity between the perpetrator and the victim. The act of killing at close quarters implies a profound connection between the individuals involved, revealing intense animosity or hatred within the familial relationship (Kleck, 2015). These findings also underscore the adverse effects associated with handguns compared to other weapons.

Perpetrator Sex and impact on weapon choice

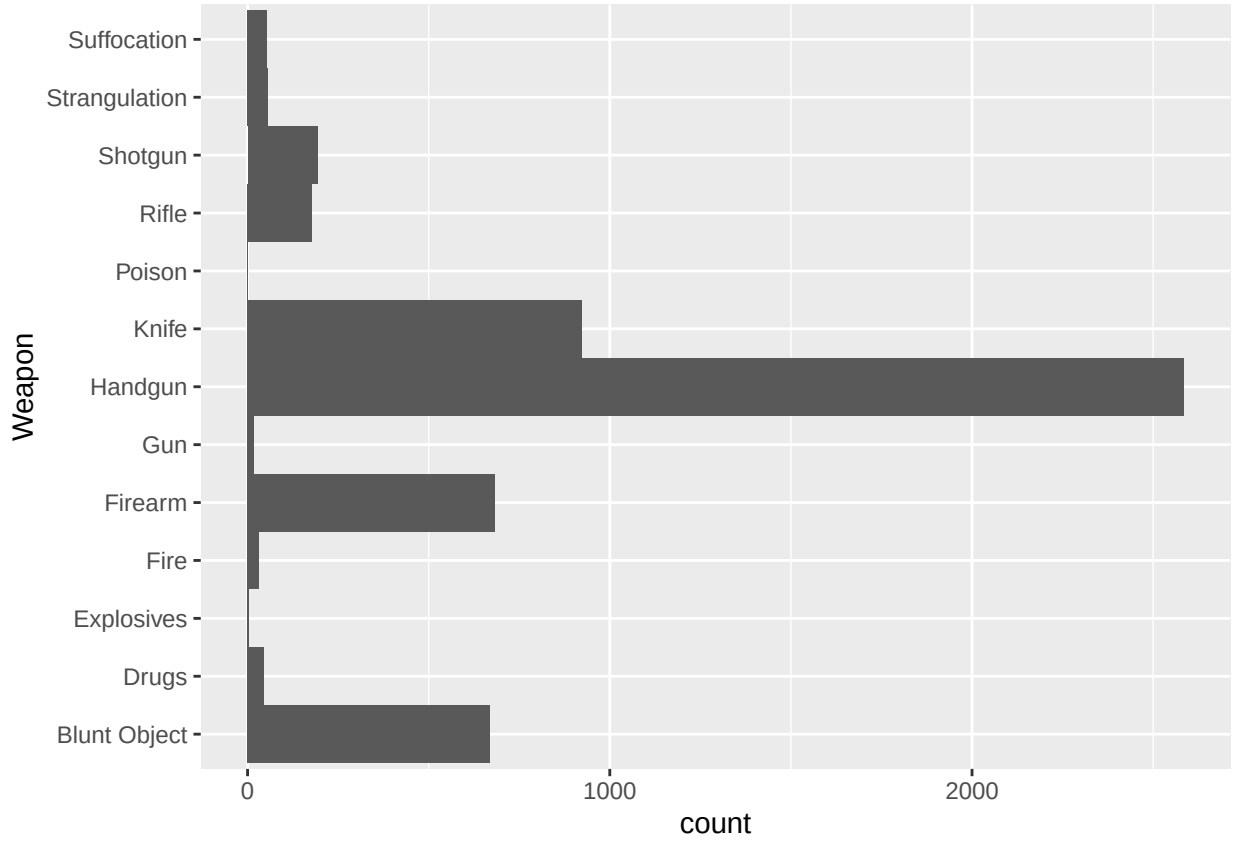
I eliminated the blank entries where no data was recorded or where the response was marked as "not sure."

In order to see for females and males, will different perpetrator sex will affect their weapon choice in homicide. Therefore, I create two seperate bar graph using ggplot function.

First graph is for male I use the filter function and pipe operator to select only rows with sex == "male"

```
## [conflicted] Will prefer dplyr::filter over any other package.
```

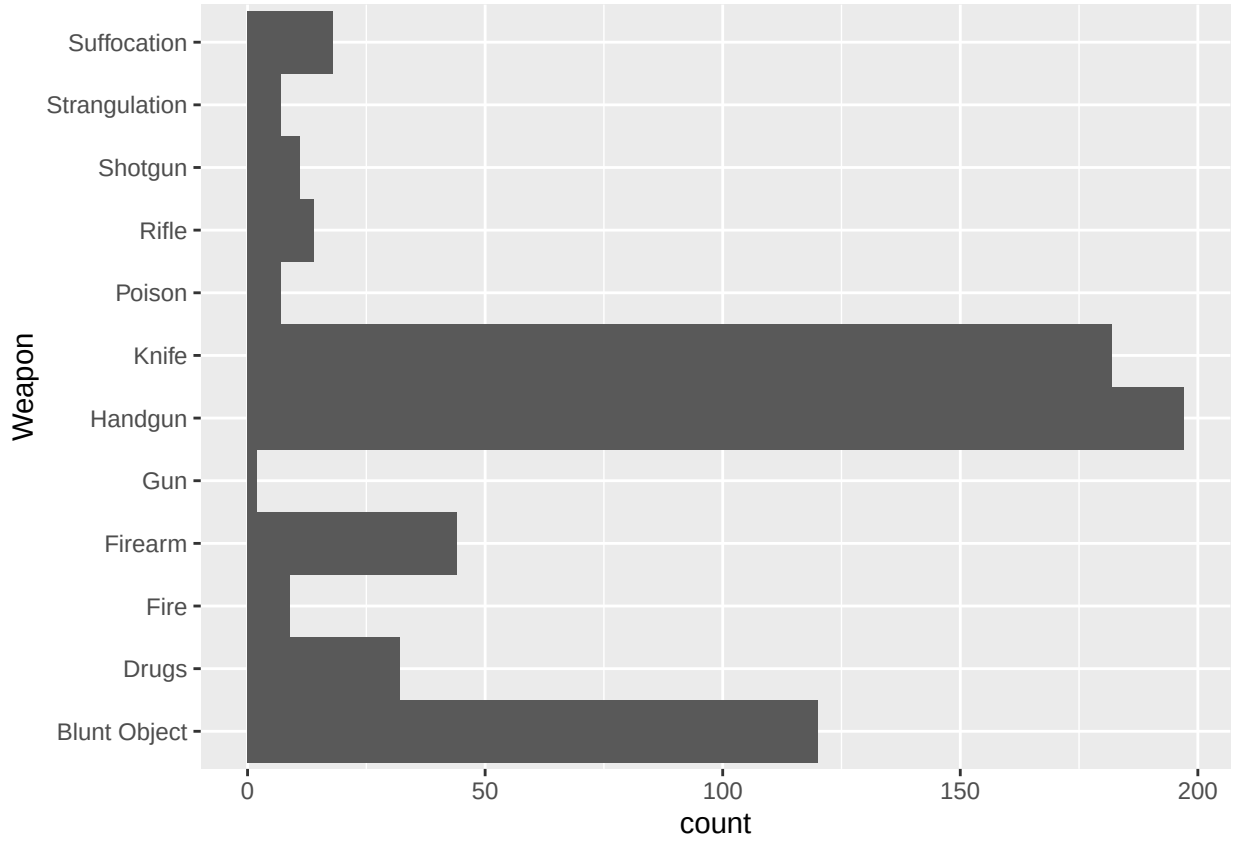
```
##   Perpetrator.Sex       Weapon
## 1            Male      Firearm
## 2            Male      Firearm
## 3            Male Blunt Object
## 4            Male      Firearm
## 5            Male        Knife
## 6            Male      Firearm
```



Based on the presented chart, the primary weapons of choice among male perpetrators in homicides are handguns, knives, and an equal distribution between firearms and blunt objects, with all three standing out prominently.

Second graph is for female I then apply the filter function and pipe operator to select only rows with sex == "female"

```
##   Perpetrator.Sex       Weapon
## 1          Female        Knife
## 2          Female  Suffocation
## 3          Female Blunt Object
## 4          Female        Knife
## 5          Female         Fire
## 6          Female      Shotgun
```



According to the chart above, female perpetrators show a tendency to use handguns, followed by knives, and blunt objects.

Overall, there is a notable absence of significant variation in weapon choices between different sexes. While both males and females commonly use knives as the second weapon of choice, females exhibit a distinctive pattern by using knives with a similar frequency to handguns. As mentioned earlier, handguns' popularity may stem from their direct impact and accessibility (STUDDERT, 2022). However, the considerably higher proportion of knife use in female-perpetrated homicides aligns with historical trends, where knives have been a prevalent choice for committing homicide. This prevalence is attributed to the ubiquitous presence of knives in households, with a significant emotional connection as kitchens, traditionally associated with women due to societal labor division, are considered a female domain (Mervosh, 2019). The connection between females, knives, and the home environment creates a sense of control, contributing to the higher frequency of knife usage in female-perpetrated homicides.

However, it is crucial to note that the overall trend might not be entirely accurate, as the majority of crimes are committed by males. The diverse contexts and motivations behind homicides for males and females contribute to the complexity of these patterns, making it challenging to establish universally representative trends.

Perpetrator Age and Weapon Type

```
##   Year Perpetrator.Age Relationship  Weapon Perpetrator.Sex RelationshipNew
## 1 2014               3       Mother Handgun            Male          Family
## 2 2014               6       Family Firearm            Male          Family
## 3 2014               9      Brother   Rifle            Male          Family
## 4 2014              10      Brother Handgun            Male          Family
## 5 2014              11      Brother Handgun            Male          Family
```

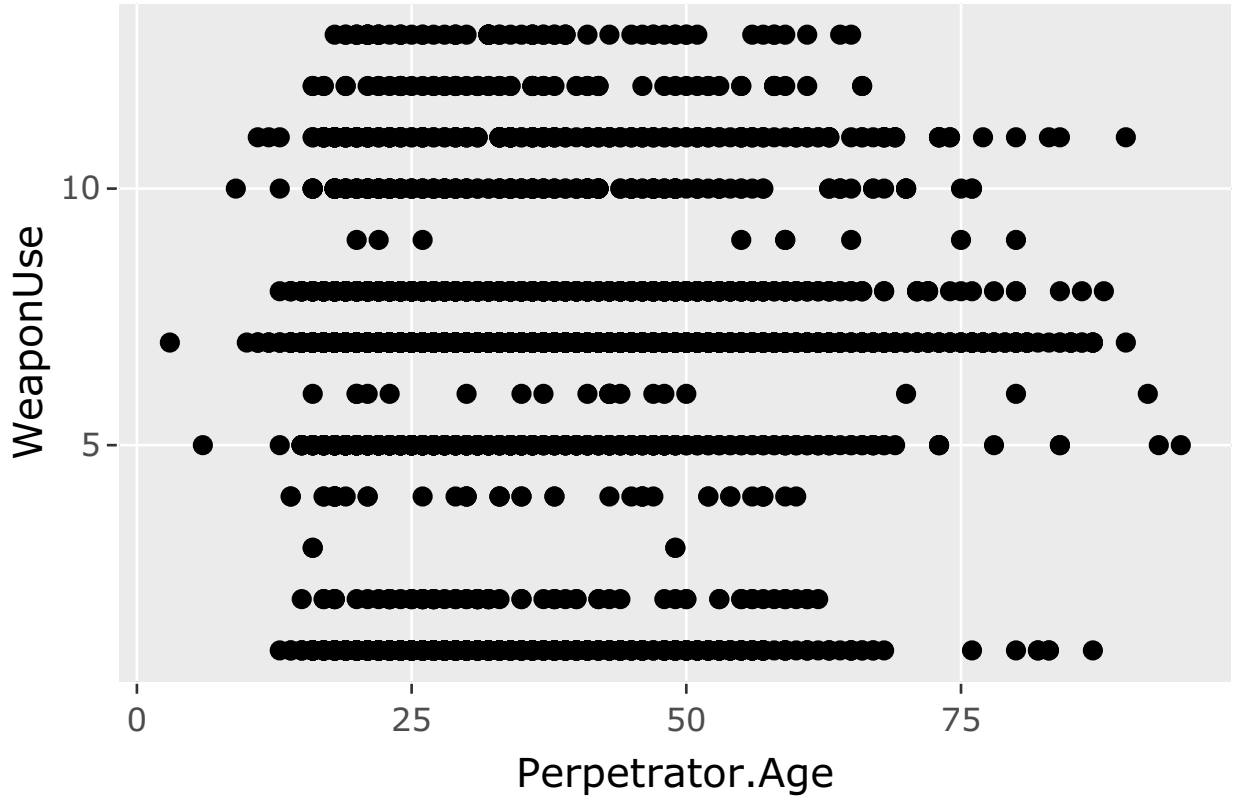
```
## 6 2014              11   Stepfather Handgun            Male     FamilyAfter
```

As the weapon use variable is categorical, I transformed it into a numeric format using the as.numeric function to facilitate the use of ggplot for graphical representation. In this case, I transform all weapon type to numeric, with "Suffocation"=1, "Strangulation"=2, "Shotgun"=3, "Rifle"=4,"Poison"=5, "Knife"=6, "Handgun"= 7, "Gun"=8, "Firearm"=9, "Fire"=10, "Explosives"=11, "Drugs"=12, "Blunt Object"=13.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   5.000   7.000   6.432   8.000  13.000
```

```
## [1] "integer"
```

To make ggolot interactive in github page, I install package called "plotly" and apply ggplotly() function



"Suffocation"=1, "Strangulation"=2, "Shotgun"=3, "Rifle"=4,"Poison"=5, "Knife"=6, "Handgun"= 7, "Gun"=8, "Firearm"=9, "Fire"=10, "Explosives"=11, "Drugs"=12, "Blunt Object"=13.

Since it is interactive, each point represent a homicide case, when you point your mouse to either one, you can see the age and weapon use for each perpetrator.

From the provided plot, it's apparent that handguns are the most commonly used weapon across all age groups, indicating consistent use regardless of age. Firearms are predominantly utilized before the age of 65. While blunt objects and knives are primarily employed before the age of 65, there is a notable cluster of homicides involving these weapons in individuals above the age of 75. The accessibility of blunt objects and knives, available in homes and shops, might explain their consistent usage across different age groups. Homicides involving rifles are more prevalent before the age of 60, potentially due to their larger size and

weight, making them less feasible for the elderly to handle. Additionally, the thrill associated with rifles might appeal more to the middle-aged and younger population.

Suffocation and strangulation show two peaks around the ages of 25 and 48. These methods may require significant physical strength and a sense of control, with younger individuals potentially acting on strong emotional impulses. The use of drugs in homicides predominantly occurs before the age of 35, with another peak observed between the ages of 53 and 63.

Among various methods, poison and explosives are the least preferred means of homicide, while fire-related incidents display a scattered distribution. The decreasing use of poison could be attributed to advancements in medical technologies. Moreover, the distribution of fire-related incidents appears scattered across different age groups.

Conclusion of main findings

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.