Computing and Visualizing Social Relationships Amongst National Homicide Trends in R

Cecilia Li

10/29/2017

Background

The Uniform Crime Reporting (UCR) Program of the FBI collects supplementary homicide data that provides information regarding the age, sex, and race of the murder victim and offender, the type of weapon used, the relationship of the victim to the offender, and the circumstance surrounding the incident. Of those factors, the social relationship between the killer and the victim pertains to the most interest above all. What constitutes of the diversity and relationship between the killer and the victim? There were a total of 17,250 reported murder and non-negligent manslaughter cases in the U.S. in 2016, according to Statista.com. Although the number of cases has declined in the past twenty years, when viewed in international comparison, the U.S. murder rate is still high. These large data sets available online have been extracted by multiple research centers and doctorals. With the use of R, our ability to faciliate rapid data analysis and visualization is realized. We now can utilize this tool with flexibility in order to adopt a conclusion or pattern which we seek to discover in this computing environment.

Motivation

As a Canadian student, coming to the United States required little effort to adapt to the new surrounding. However, the only arising issue that resonates with me strongly is the safety here. My hometown, Toronto, was a place where I would feel safe walking alone in alleyways at 2AM in the morning. Coming to Berekely truly changed my comfortability in doing so, and raised my awareness in not only campus safety, but moreoever the larger scope of safety and homicide rates in this country. Recent protests and gunshoots, along with numerous riots that were targeted towards minority groups, made me ponder about the relationship between people involved in those killings. I believe that the differentiating factor between humans and other animals are emotions. Emotions play a huge factor in our actions and decisions. Therefore, I decided to look into how relationships play a role in US homicide, particularly through the perpetrator and victim's perspective and its association with race and age.

memecenter.com MemeCenter

Introduction

This post will be about applying R functions to extract trends from the large data set. The central theme amongst all visualizations will lead to further knowledge about the different attributes of perpetrators and victims in terms of age and race, as well as the most common victim-murderer reltaionships. Due to the nature of the raw data, codes and functions will be more complex and complicated when expressed in a cell. I will be making comments throughout the way about the significance of the visualization. At the end, I will be drawing conclusions from the trends on a macro scale. Hopefully, you will be able to learn something new about the social and ethnical relationships involved in murder!

A Few Things To Keep In Mind:

Main Goal

The main goal of this post is to extrapolate data, visualize graphs and draw assumptions and trends using R on the topic surrounding relationships and people in homicide in the United States.

Data Used

- In order to locate appropriate and useful data, I conducted research on the web and found data source with useful datasets for our analysis. All the data tables were downloaded from a website called Kaggle, which is useful for searching and downloading datasets. The dataset includes murders from FBI's U.S homicide reports ranging from 1980 to 2014.

Murder Accountability - Homicide Reports

Examples of Analysis

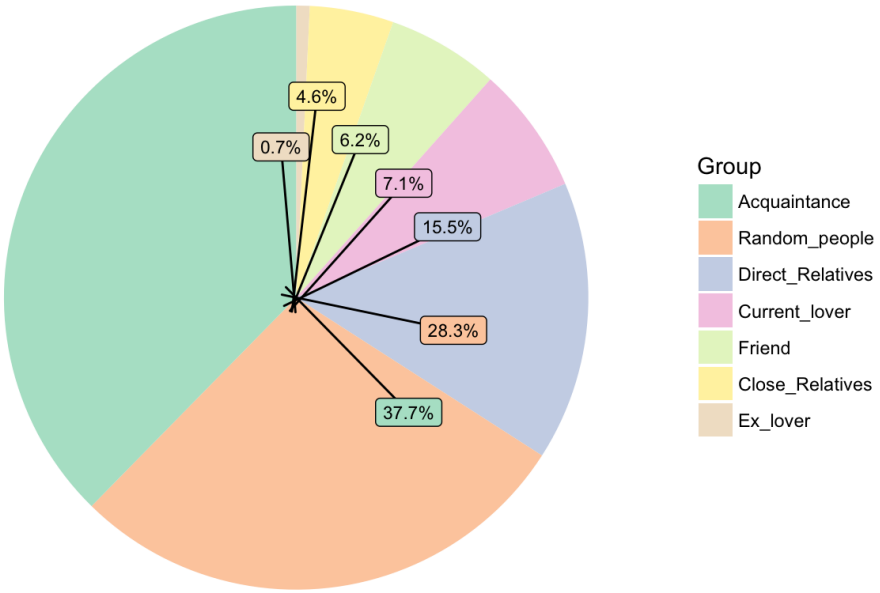
1. Loading the relationship data

```
library(DataComputing)
library(dplyr)
library(ggplot2)
library(ggrepel)
library(forcats)
library(scales)
library(RColorBrewer)
library(wesanderson)
library(DT)
library(xtable)
library(knitr)
Cleandata<- read.csv(file="~/Desktop/cleanrelationship.csv")
```

2. Compute the proportion of roles as victim in relative to the killer

```
Direct_Relatives <- c("Wife","Husband","Brother","Sister","Son","Daughter","Common_Law Wife","Common_law husband"
)

Close_Relatives<- c("Stepson","Stepdaughter","Stepmom","Stepfather","In-Law","Family")

Ex_lover<- c("Ex-Husband","Ex-Wife")
Current_lover<- c("Girlfriend","Boyfriend","Boyfriend/Girlfriend")

Acquaintance<-c("Acquaintance","Neighbor","Employer","Employee")
Friend<- c("Friend")
Random_people <- c("Stranger")
Cleandata$Group<- with(Cleandata,ifelse(Relationship %in% Direct_Relatives,"Direct_Relatives",ifelse(Relationship
%in% Close_Relatives, "Close_Relatives",ifelse(Relationship %in% Close_Relatives,"Close_Relatives",ifelse(Relatio
nship%in%Current_lover,"Current_lover",ifelse(Relationship%in% Ex_lover,"Ex_lover",ifelse(Relationship%in% Acquai
ntance,"Acquaintance",ifelse(Relationship%in%Friend,"Friend","Random_people")))))))
```

```
library(dplyr)
library(ggplot2)
library(ggrepel)
library(forcats)
library(scales)
library(RColorBrewer)
RelationshipGroup<-Cleandata%>%group_by(Group)%>%summarise(Count=n())%>%arrange(desc(Count)) %>%mutate(proportion
=percent(Count/sum(Count)))

blank_theme <- theme_minimal()+
  theme(
  axis.title.x = element_blank(),
  axis.title.y = element_blank(),
  panel.border = element_blank(),
  panel.grid=element_blank(),
  axis.ticks = element_blank(),
  plot.title=element_text(size=14, face="bold")
  )

RelationshipGroup%>%ggplot(aes(x="",y=Count,fill=fct_inorder(Group)))+  geom_bar(width = 1, stat = "identity")+co
ord_polar("y", start=0)+scale_fill_brewer(palette="Pastel2") + blank_theme+theme(axis.text.x=element_blank())+geo
m_label_repel(aes(label = proportion), size=3, show.legend = F, x=0.45,nudge_x =0.6,nudge_y = 0.2)+guides(fill =
guide_legend(title = "Group"))
```



As shown in the pie chart, the group that has the largest share in the pie chart is “acquaintance”, 37.7%, followed by “random people”, 28.3%. However, as a statistical point of view, this representation is not as effective if we want to compare the absolute values as well as want to see more distinguishable results amongst the categories. Furthermore, going in retrospect, let’s break down each of the categories further:

- let “Direct_Relatives” contain “Wife”, “Husband”, “Brother”, “Sister”, “Son”, “Daughter”, “Common_Law Wife”, and “Common_law husband”
- let “Close_Relatives” contain “Stepson”, “Stepdaughter”, “Stepmom”, “Stepfather”, “In-Law”, and “Family”;
- let “Ex_lover” contain “Ex-Husband” and “Ex-Wife”;
- let “Current_lover” contain “Girlfriend”,“Boyfriend” and “Boyfriend/Girlfriend”
- let “Acquaintance” contain “Acquaintance”, “Neighbor”, “Employer” and “Employee”
- let “Friend” just be “Friend”
- let the rest be classified as “Random_people”

3. Visualizing in alternative graphs and details

```
kable(RelationshipGroup,caption="Relationship groups in homocide")
```

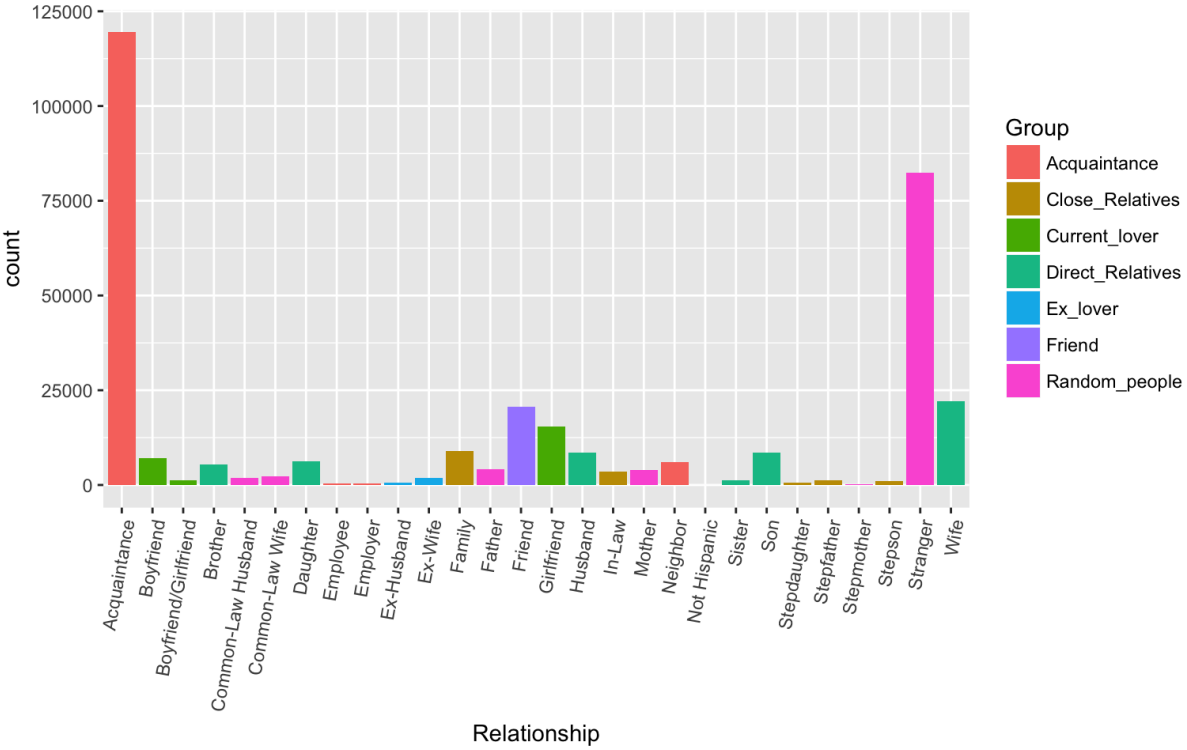
Relationship groups in homocide

Group	Count	proportion
Acquaintance	126350	37.7%
Random_people	94827	28.3%
Direct_Relatives	51882	15.5%
Current_lover	23683	7.1%
Friend	20742	6.2%
Close_Relatives	15592	4.6%
Ex_lover	2494	0.7%

Murder-Victim Relationship Histogram

```
Cleandata%>%ggplot(aes(x=Relationship,fill=Group))+geom_histogram(stat="count")+theme(axis.text.x = element_text(
angle = 80, hjust = 1))
```

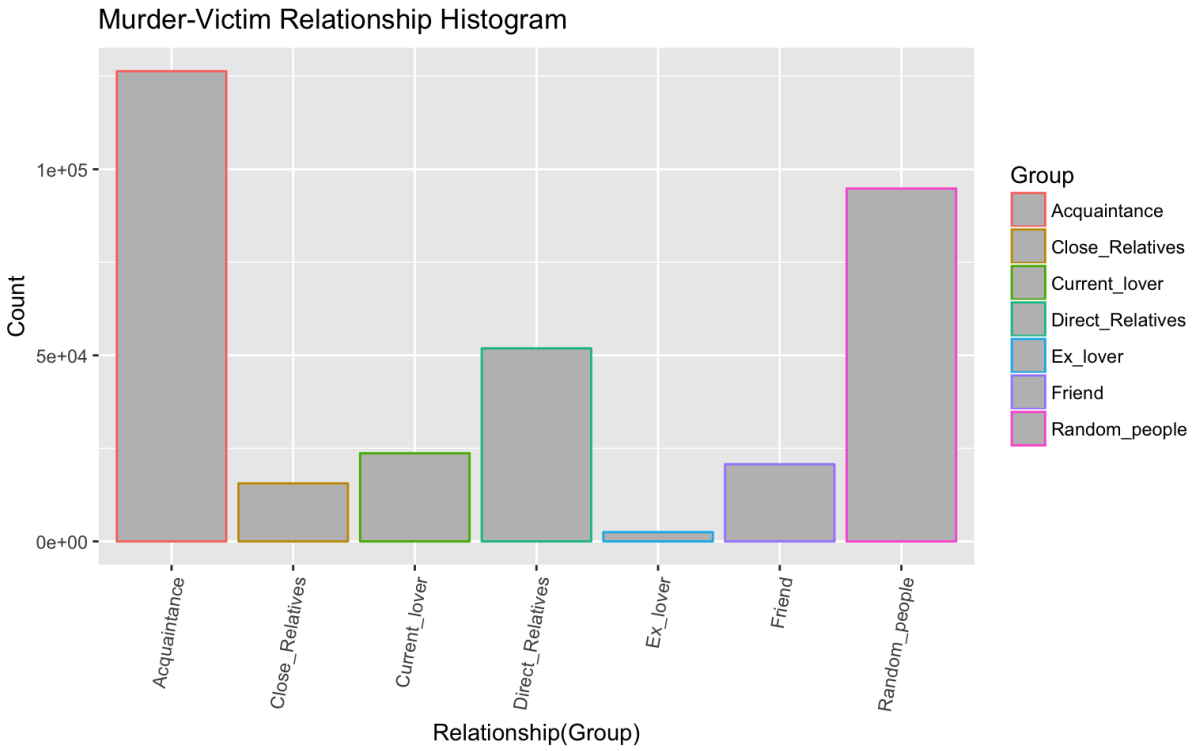
```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



Now there are 28 types of relationships. All of them clearly illustrate the frequencies of relationship types between the murderer and the victim.

```
Cleandata%>%ggplot(aes(x=Group,col=Group))+geom_histogram(stat="count",fill="grey")+theme(axis.text.x = element_t
ext(angle = 80, hjust = 1))+labs(x="Relationship(Group)",y="Count",title="Murder-Victim Relationship Histogram")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

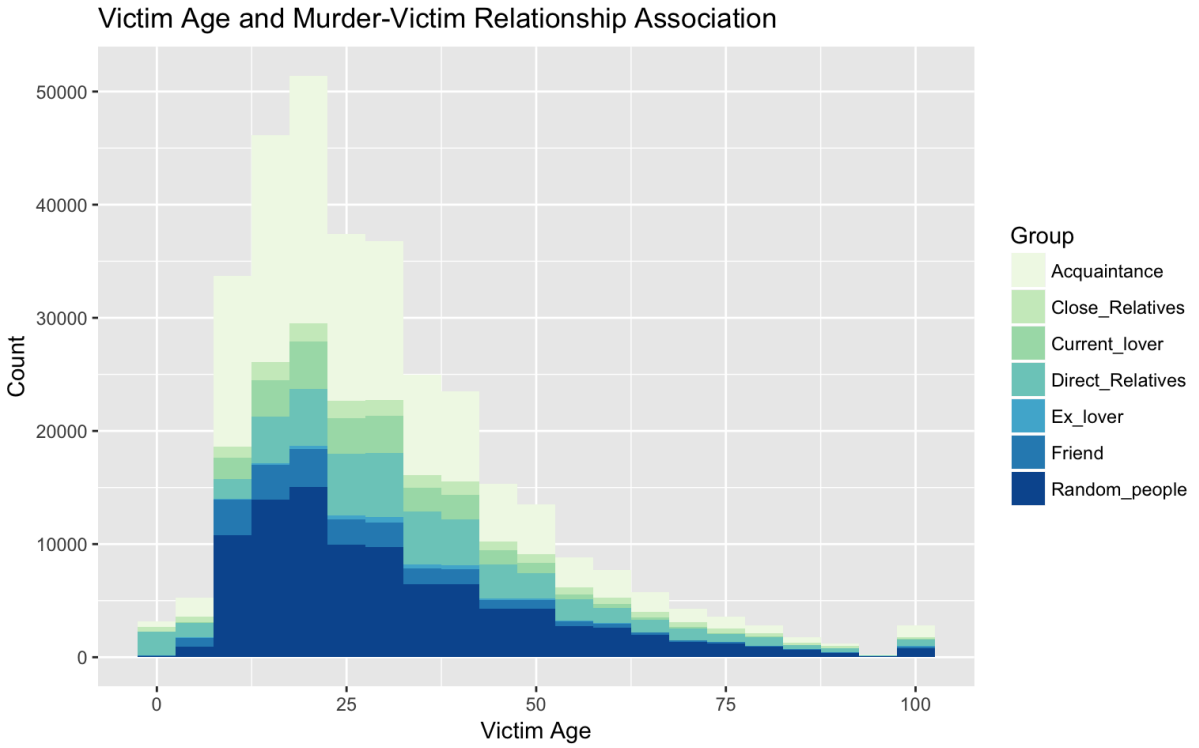


Drawing the above graphs conclusively, I see that the results are quite interesting. As aforementioned, the pie chart illustrates the largest portion being “acquaintance”. This is very contradictory to common assumptions, as we may not expect that people get murdered by someone they knew, including relatives, lovers, and acquaintances. I am genuinely surprised to discover that out of all possibilites, 15.5% of the killing occured between “relatives”. Furthermore, the fact that 7.1% of all victims were in a love relationship with the person whom killed them is even scarier!

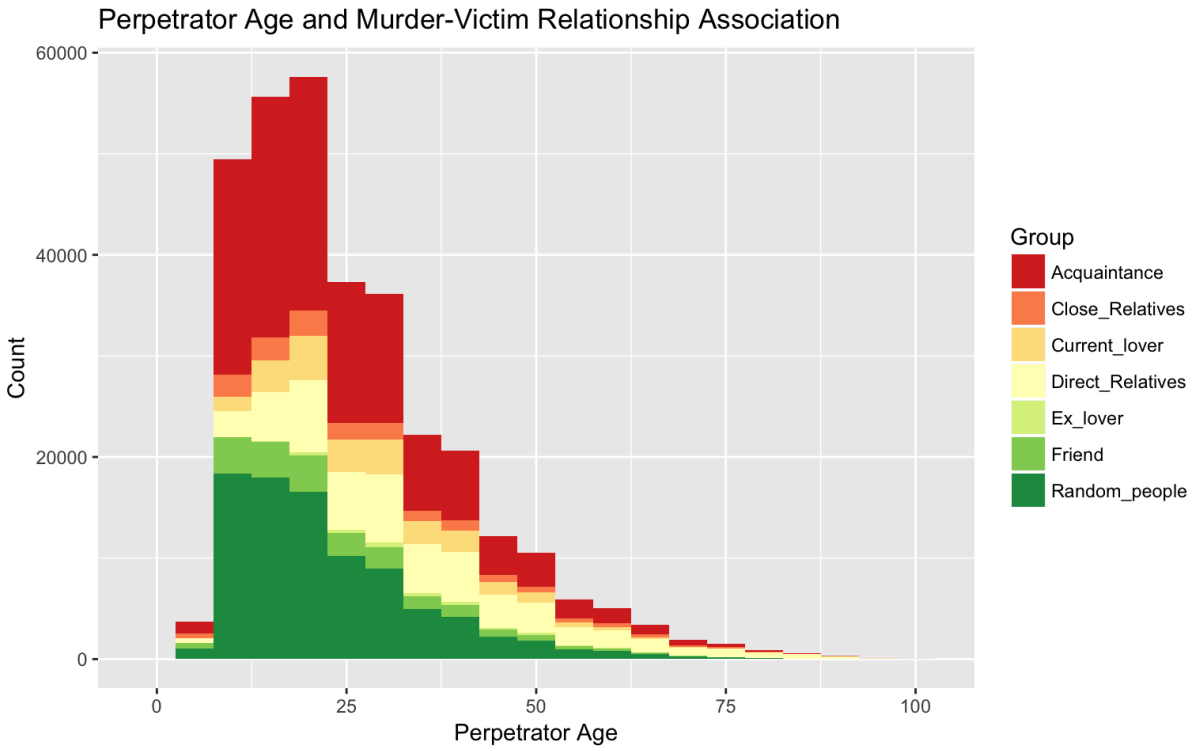
3. Looking through the scope of time

I wanted to look at how these relationships may differ or be impacted by the factor of age. This may be significant, as mental development plays a significant role in the societal trends and impulsions in killing. The age of the victim may also be an element that accounts for the killer’s target.

```
Cleandata%>%filter(Victim.Age!=0)%>%ggplot(aes(x=as.numeric(Victim.Age),fill=Group))+geom_histogram(binwidth=5)+l
abs(x="Victim Age",y="Count",title="Victim Age and Murder-Victim Relationship Association")+scale_fill_brewer(pal
ette="GnBu")
```




```
Cleandata%>%filter(Perpetrator.Age!=0)%>%ggplot(aes(x=as.numeric(Perpetrator.Age),fill=Group))+geom_histogram(bin
width=5)+labs(x="Perpetrator Age",y="Count",title="Perpetrator Age and Murder-Victim Relationship Association")+s
cale_fill_brewer(palette="RdYlGn")
```



The results are interesting. Let's dive in on how age plays a role in terms of the relationship groups from the perspective of the victim and the perpetrator. Both graphs peak in the range between 12-20 years old and gradually skew to the right. This might be because teenagers are more emotional and have a higher tendency to do stupid things…

If we look at the perpetrator age and association graph, some interesting phenomenons appear: Between age 10-20, there is an increase in total counts of homicide, but the proportions of relationships being "random" decreases, with a direct increase in "direct relatives" proportion. And if we look at the victim graph, there are outliers at the very end with Age 100. This is supported by the Law of Jungle: The Weak are the prey of the Strong. People may think that older people would be a burden to them and, hence, triggers their motive to kill.

4. Let's talk about diversity

Lastly, I want to look into the role race plays in the relationships of the murder trends.The table below shows in each perpetrator ethnicity group, what kind of relationship do they target the most frequent and the least frequent.

in each murder ethnicity, what kind of relationship mostly go for

```
Race_group_data<-Cleandata%>%filter(Perpetrator.Race!=24)%>%filter(Perpetrator.Race!=25)%>%group_by(Perpetrator.Race,Group)%>%summarise(total=n())

kable(Race_group_data[!duplicated(Race_group_data$Perpetrator.Race),],caption="In each murder ethnicity, what kind of relationship mostly go for")
```

In each murder ethnicity, what kind of relationship mostly go for

Perpetrator.Race	Group	total
Asian/Pacific Islander	Acquaintance	1279
Black	Acquaintance	69145
Native American/Alaska Native	Acquaintance	1203
White	Acquaintance	54723

in each murder ethnicity, what kind of relationship leastly go for

```
Least_Race_group_data<-Cleandata%>%filter(Perpetrator.Race!=24)%>%filter(Perpetrator.Race!=25)%>%group_by(Perpetrator.Race,Group)%>%summarise(total=n())%>%arrange(Perpetrator.Race,total)

kable(Least_Race_group_data[!duplicated(Least_Race_group_data$Perpetrator.Race),],caption="In each murder ethnicity, what kind of relationship leastly go for")
```

In each murder ethnicity, what kind of relationship leastly go for

Perpetrator.Race	Group	total
Asian/Pacific Islander	Ex_lover	45
Black	Ex_lover	597
Native American/Alaska Native	Ex_lover	23
White	Ex_lover	1829

```
Race_relationship_data<-Cleandata%>%filter(Perpetrator.Race!=24)%>%filter(Perpetrator.Race!=25)%>%group_by(Perpetrator.Race,Relationship)%>%summarise(total=n())
kable(Race_relationship_data[!duplicated(Race_relationship_data$Perpetrator.Race),],caption="the kinds of relationship that has most murder rate from each perpetrator's race perspective")
```

the kinds of relationship that has most murder rate from each perpetrator's race perspective

Perpetrator.Race	Relationship	total
Asian/Pacific Islander	Acquaintance	1192
Black	Acquaintance	66799
Native American/Alaska Native	Acquaintance	1154
White	Acquaintance	50336

```
Least_Race_Relationship_data<-Cleandata%>%filter(Perpetrator.Race!=24)%>%filter(Perpetrator.Race!=25)%>%group_by(Perpetrator.Race,Relationship)%>%summarise(total=n())%>%arrange(Perpetrator.Race,total)

kable(Least_Race_Relationship_data[!duplicated(Least_Race_Relationship_data$Perpetrator.Race),],caption="the kinds of relationship that has least murder rate from each perpetrator's race perspective")
```

the kinds of relationship that has least murder rate from each perpetrator's race perspective

Perpetrator.Race	Relationship	total
Asian/Pacific Islander	Common-Law Husband	3
Black	Stepmother	39
Native American/Alaska Native	Employee	1
White	Stepmother	156

Without even plotting this, it is evident that the pattern is sporadic, and that race does not play a role in the murder - victim relationship. This is very surprising, as many people would presume otherwise due to the historically overwhelming number of protests targeted against a group of race. For instance, the Mass Racial Violence involves discrimination and brutality towards African American, Native American, and Latin American, and, sadly much of the inequality still prevails today. However, it seems like the national data indicates otherwise on a sociological perspective of such incidences.

Conclusion

Generally, relationship plays an important role from the perspective of the killer, or perpetrator. Surprisingly, race does not play a role in the types of relationships between the killer and the victim in a homicide incident. As shown in the data, direct relatives and even lovers are sometimes targeted whereas acquaintances and random people occupy larger proportions. Age is also a determinant in the relationships amongst people, and this is supported not only through data analysis conducted above, but also various social-psychological theories of age as well as the Law of Jungle.

Blogger's Takeaway Message

It is always easy to notice patterns in societies, but never easy to have the persistency to go above and beyond to conquer that curiosity. What is amazing about the project is the experience of having been able to not only explore an area of interest, but moreover apply the skillsets in R to add significance to the data bulks there are online. A major role a data scientist plays is the ability to identify the problem, to sort out the data in ways that specifically target that problem, and to ultimately assess the applicable skills on hands and resolve the problem. The two most powerful skillsets are being able to filter out unnecessary data, and being able to zoom in and target different but converging perspectives.

References

https://ucr.fbi.gov/crime-in-the-u.s/2011/crime-in-the-u.s.-2011/offenses-known-to-law-enforcement/expanded/expanded-homicide-data

https://www.statista.com/statistics/195331/number-of-murders-in-the-us-by-state/

Cooper, Alexia, and Erica L. Smith. "Bureau of Justice Statistics Home Page." Bureau of Justice Statistics (BJS). BJS Statisticians, 1 Nov. 2011. Web. 03 May 2017.

Levitt, Steven D. "Understanding Why Crime Fell in the 1990s: Four Factors That Explain the Decline and Six That Do Not." Journal of Economic Perspectives 18.1 (2004): n. pag. Web.

https://www.kaggle.com/murderaccountability/homicide-reports

http://dictionary.cambridge.org/dictionary/english/law-of-the-jungle

https://en.wikipedia.org/wiki/Mass_racial_violence_in_the_United_States

Final Words

Thank you for reading until the end. I hope you enjoyed learning about the relationship trends in United States homicide incidences as much as I enjoyed researching, visualizing, and analyzing this topic!


dont joke about murder i was
murdered once and it offends me