Fatal Police Shootings

Exploratory Data Analysis

The datacan be broadly classified into two groups: victim descriptors and situation sescriptors. Victim descriptors provide information about the victim that is independent of the specific altercation that resulted in death. Situation descriptors describe the immediate environment of the incident and the interaction between the involved parties, as described by the 'Armed' and 'Classification' fields.

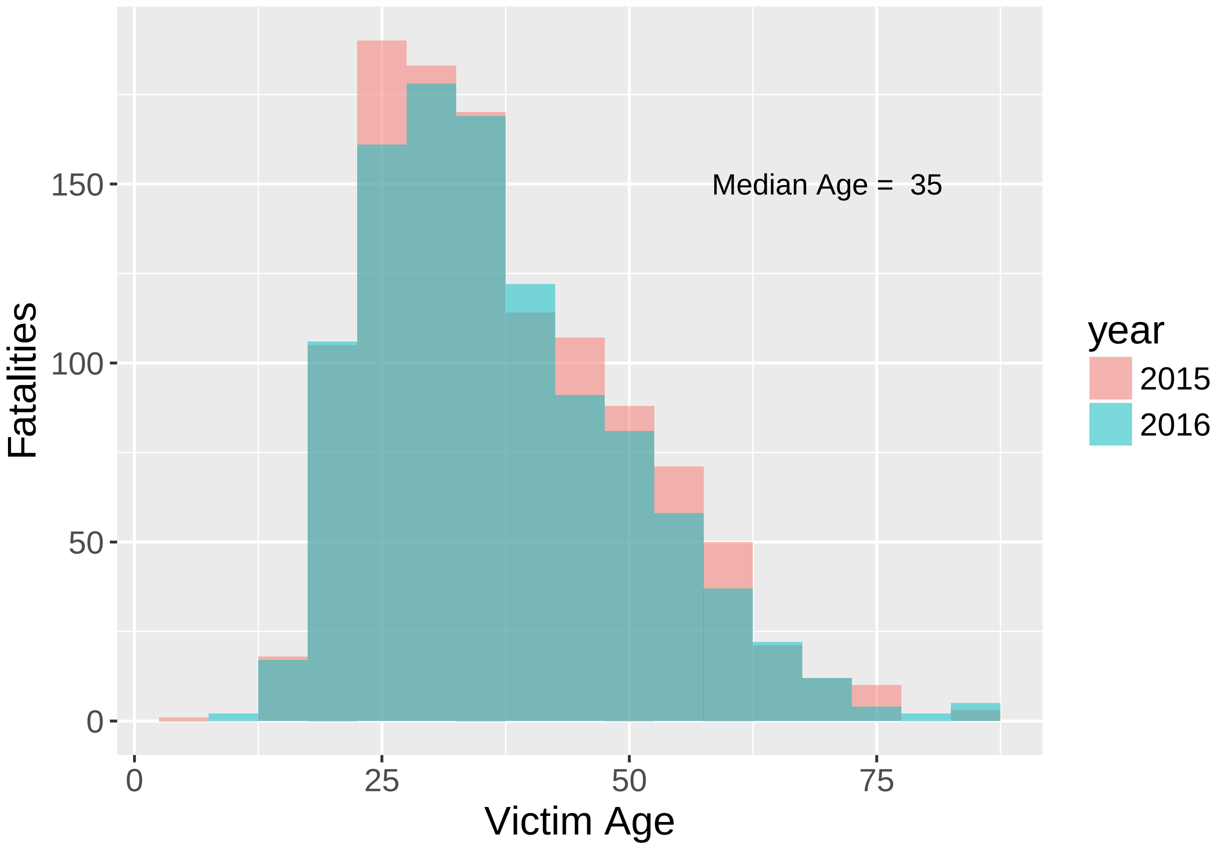
Splitting the descriptors into these groups is useful because the victim and situation information answer the questions "What people are most frequently victims of shootings?" and "What situations result in police gun shooting fatalities?", respectively.

Victim Descriptors	*Situation Descriptors*
Name	Date
Age	Location (State & City)
Race	Classification (Manner of Death)
Gender	Armed (Type of Weapon on Victim)

Victim Data

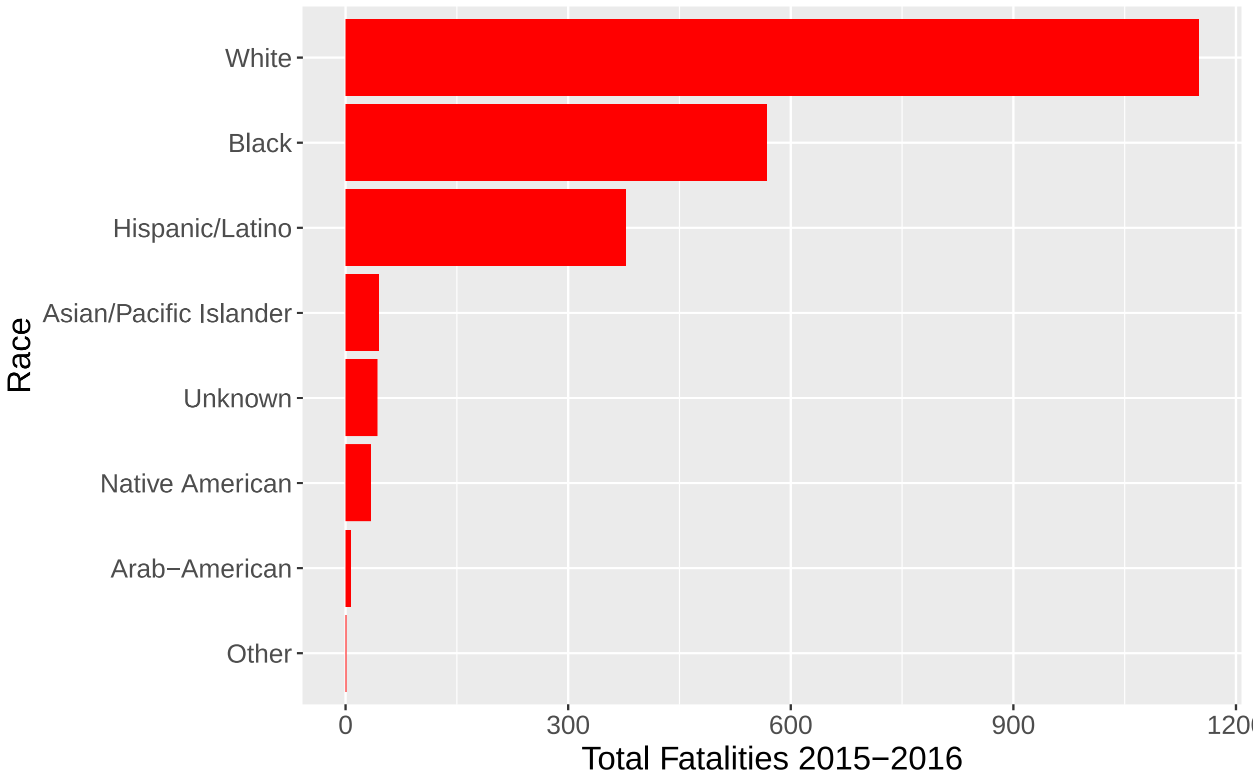
The age distribution of police-related fatalities is nearly identical for both 2015 and 2016 with a median value of 35 years. The tail at the higher ages indicates that the chance of police altercations becoming lethal reduces gradually as people become older. This may be due to a decreased inclination for police to use lethal force or a decreased ability for the person to present himself as a significant threat.

```
##Victim

    #Age Distribution
    g.guard.age<-ggplot(data = clean.guard.dt[age>=0],aes(x=age,fill=year))+
        geom_histogram(position='identity',alpha=0.5,binwidth = 5)+
        xlab('Victim Age')+ylab('Fatalities')+theme(text = element_text(size = 15))+
        annotate("text", x = 70, y = 150,
                 label = paste('Median Age = ',median(clean.guard.dt[age>=0]$age)))

    g.guard.age
```



The race of an individual appears to be a significant indicator of police-related fatalities. There is a clear disparity with 3 of the 8 reported races accounting for 94.2% of all reported fatalities. It should be noted that the relative values of the reported need to be adjusted for population before the effects of race on odds of police-related fatality can be estimated.

```
#Race Distribution
g.guard.race<-ggplot(data=clean.guard.dt, aes(x=reorder(race, table(race)[race]))) +
    geom_bar(fill='red')+
    xlab('Race')+ylab('Total Fatalities 2015-2016')+
    theme(text = element_text(size = 15))+
    coord_flip()

g.guard.race
```




```
#Race ECDF
guard.tbl.race<- table(clean.guard.dt$race)
guard.tbl.race<-guard.tbl.race[order(guard.tbl.race)]
guard.race.ECDF<-tbl.ECDF(guard.tbl.race) #ECDF of WP weapons

print(paste0('White, Black, & Hispanic Deaths Account for ',
                    as.character(round(guard.race.ECDF[3]*100,digits=1)),
                        '% of all fatalities' ))
```

```
## [1] "White, Black, & Hispanic Deaths Account for 94.2% of all fatalities"
```

Of the 3 victim descriptors, gender is the most significant predictor with men being 20 times more likely of being killed by police than women.

```
clean.guard.dt[,.N,by=gender]
```

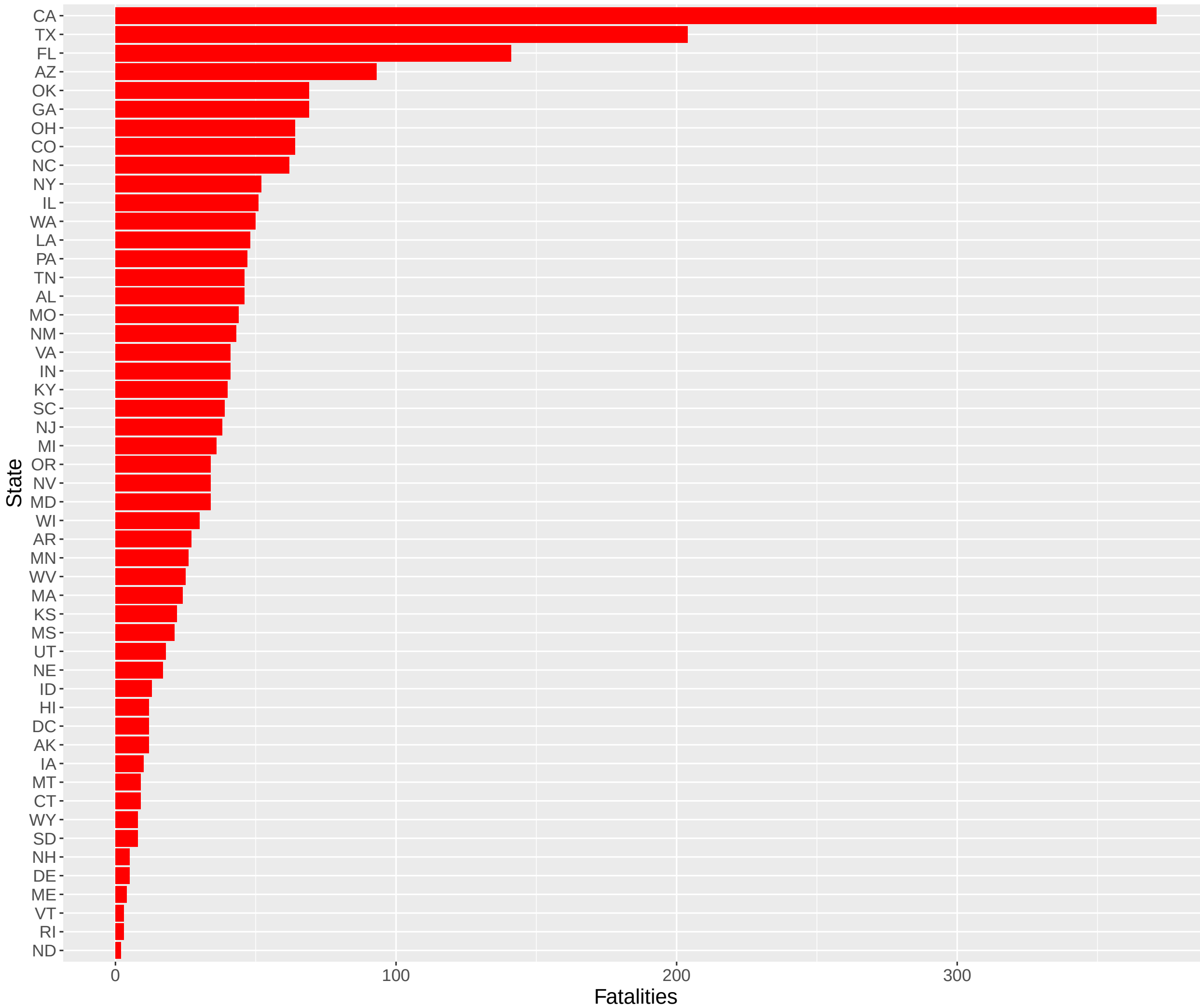
```
##            gender    N
## 1:           Male 2115
## 2:         Female  110
## 3: Non-conforming    1
```

Situation Data

The distribution of fatalities across the United States indicates that there is a much higher occurrence of police-related deaths in California, Texas, Florida, and Arizona compared to the rest of the states. If the fatalities in these top 4 states are ignored, the number of deaths decreases slowly from across states implying that police-related deaths occur at the same rate per year regardless of location.

```
##Situation
    #Location (States)
    g.guard.state<-ggplot(data=clean.guard.dt,
```

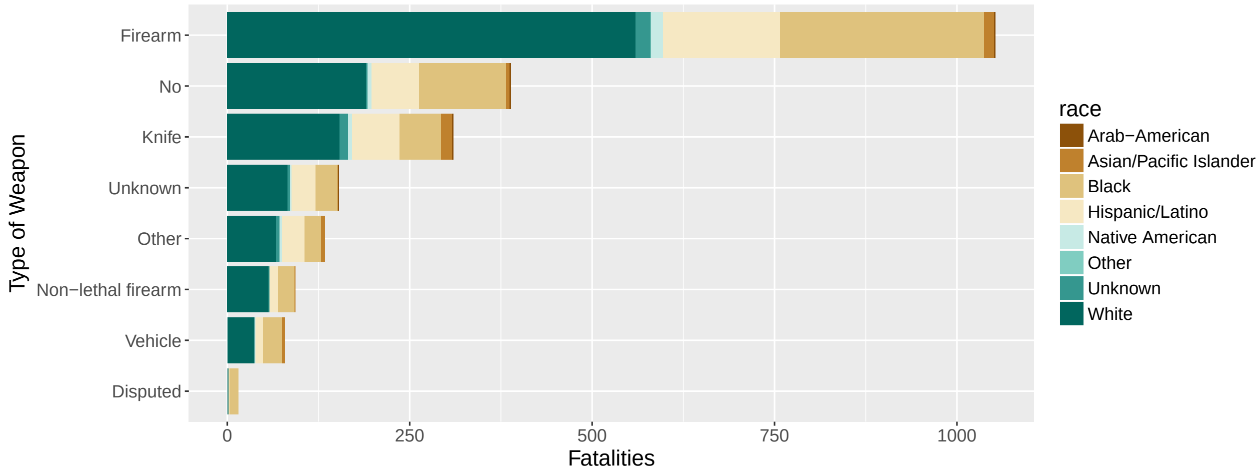
```
                     aes(x=reorder(state, table(state)[state])) +
    geom_bar(fill='red')+coord_flip()+
    xlab('State')+ylab('Fatalities')+
    theme(text = element_text(size = 15))

g.guard.state
```



A plot of the type of weapon in possession by victims is intriguing. It makes sense that the weapon that is most frequently associated with police using lethal force is a firearm. After firearms however, most fatalities occur when people are unarmed. This is important because it means that even when a person does not have a weapon, there exist conditions when police will deem it necessary to use lethal force.

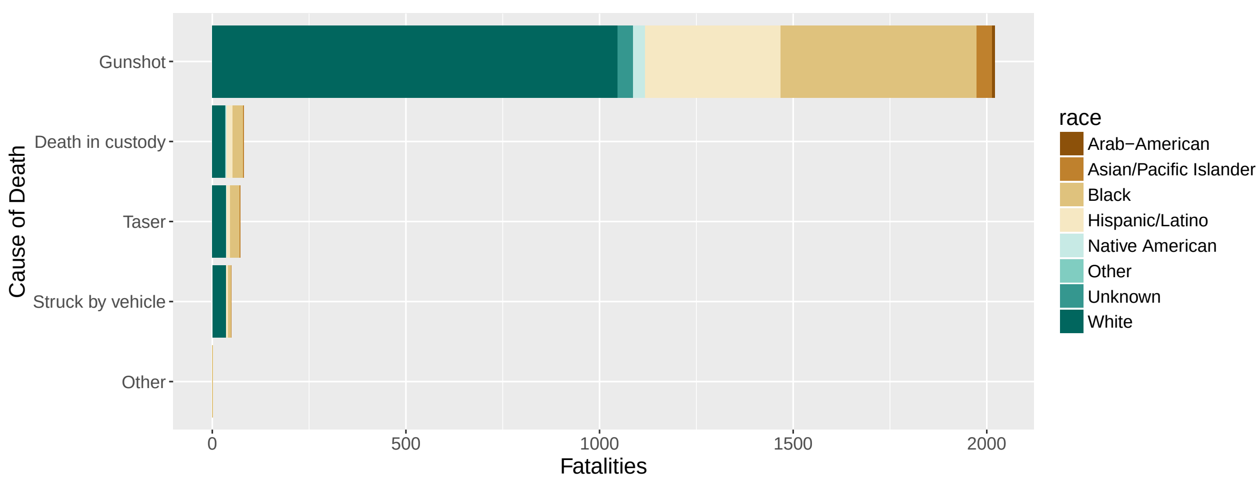
```
#Armed
g.guard.armed<-ggplot(data=clean.guard.dt,
                     aes(x=reorder(armed, table(armed)[armed]),fill=race)) +
    geom_bar()+ scale_fill_brewer(palette = 'BrBG')+
    coord_flip()+
    xlab('Type of Weapon')+ylab('Fatalities')+
    theme(text = element_text(size = 15))
```

```
g.guard.armed
```



The method used by police that most frequently results in a fatality is gunshot. Of the 2226 documented fatalities, 2021 were the result of firearm use by the police. All analyses should be done in the context of gunshot fatalities because those deaths essentially describe the entire dataset with the other classifications providing relatively little information.

```
#Cause of Death
   g.guard.COD<-ggplot(data=clean.guard.dt,
                      aes(x=reorder(classification,
                                    table(classification)[classification]),fill=race)) +
     geom_bar()+coord_flip()+
     xlab('Cause of Death')+ylab('Fatalities')+
     scale_fill_brewer(palette = 'BrBG')+
     theme(text = element_text(size = 15))

   g.guard.COD
```



Analysis of Descriptor Interactions

Direct analysis of the victim and situation descriptors provides a coarse view of the possible predictors of police-related fatalities. They are not useful in isolation because when a police officer decides to use lethal force, he must consider multiple conditions that are occurring simultaneously. An example would be 'Is my attacker physically strong and does he possess a weapon?'.

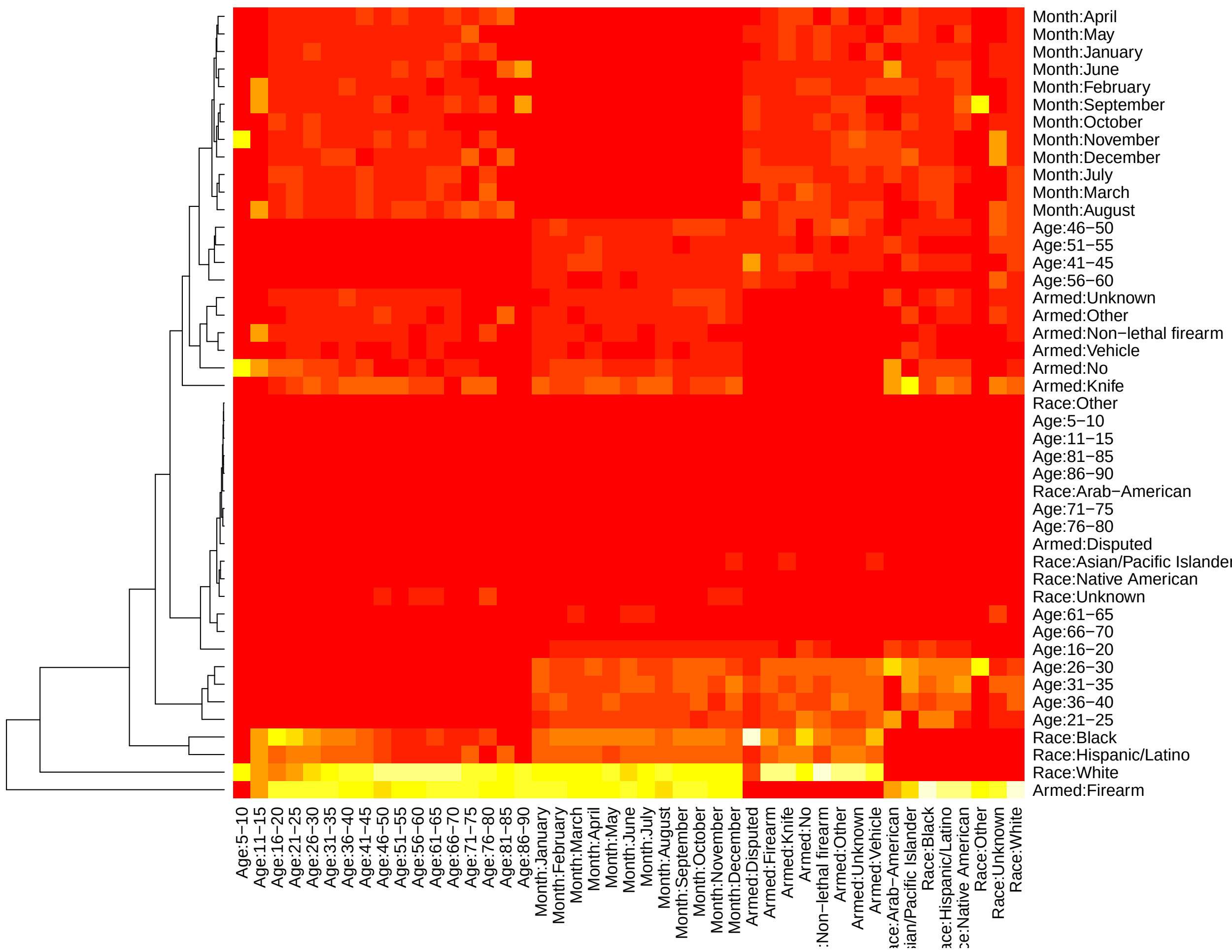
Since most of the variables in the dataset are not numeric, a standard correlation plot cannot be used to identify multiple conditions that describe a single fatality. However, it can be analyzed using a heatmap that indicates how many times pairs of conditions occur. The presented heatmap indicates this for male gunshot victims. Victim age has been binned into 5 year increments. Red and yellow regions indicate low and high occurrence of the condition pairs, respectively.

Variables of Interest: Age, Month, Type of Weapon, & Race

```
cond.target<-function(field, target){
    #(target-target%%5)/5 + 1

    age.levels <- as.character(c(1:17))

    month.levels <- c(" 'January' " ," 'February' ", " 'March' "," 'April' "," 'May' ", " 'June' ", " '
                        " 'August' "," 'September' ", " 'October' ", " 'November' ", " 'December' ")

    armed.levels<-c( " 'Disputed' " ," 'Firearm' "," 'Knife' "," 'No' ",
                     " 'Non-lethal firearm' "," 'Other' "," 'Unknown' "," 'Vehicle' ")

    race.levels<-c(" 'Arab-American' "," 'Asian/Pacific Islander' ", " 'Black' "," 'Hispanic/Latino' ",
                    " 'Native American' "," 'Other' "," 'Unknown' "," 'White' " )

    v.str <- ifelse(field==1,'((age - age%%5)/5)',
                ifelse(field==2,'month',
                       ifelse(field==3,'armed','race')))

    trgt.str<-ifelse(field==1,age.levels[target],
                ifelse(field==2,month.levels[target],
                       ifelse(field==3,armed.levels[target],race.levels[target])))

    paste0(v.str,'==',trgt.str)
} #Creates Conditional Indexing String


#v('gunshot' & 'male') = [ [age], [month] ,[armed], [race] ]
#age.levels <- as.character(c(1:17))
age.levels <- c('Age:5-10','Age:11-15','Age:16-20',
                'Age:21-25','Age:26-30','Age:31-35','Age:36-40',
                'Age:41-45','Age:46-50','Age:51-55','Age:56-60',
                'Age:61-65','Age:66-70','Age:71-75','Age:76-80',
                'Age:81-85','Age:86-90')

month.levels <- c( 'Month:January'  , 'Month:February' ,  'Month:March' , 'Month:April' , 'Month:May' ,
                   'Month:June' ,  'Month:July' ,  'Month:August' , 'Month:September' ,
                   'Month:October' ,  'Month:November' , 'Month:December' )

armed.levels<-c(  'Armed:Disputed'  , 'Armed:Firearm' , 'Armed:Knife' , 'Armed:No' ,
                  'Armed:Non-lethal firearm' , 'Armed:Other' , 'Armed:Unknown' , 'Armed:Vehicle' )

race.levels<-c( 'Race:Arab-American' , 'Race:Asian/Pacific Islander' ,  'Race:Black' ,
                'Race:Hispanic/Latino' ,  'Race:Native American' ,
                'Race:Other' , 'Race:Unknown' , 'Race:White'  )
```

```r
pred.names<-c(age.levels,month.levels,armed.levels,race.levels)

age.bins.N <-length(age.levels)
month.N <- length(month.levels)
armed.N <- length(armed.levels)
race.N<-length(race.levels)
total.levels<-age.bins.N+month.N+armed.N+race.N

mat.heat<-matrix(nrow=total.levels,ncol=total.levels)

for (i in 1:total.levels) {
    cond1.field<-ifelse(i<age.bins.N+1,1,
                ifelse(i<age.bins.N+month.N+1,2,
                    ifelse(i<age.bins.N+month.N+armed.N+1,3,4)))

    cond1.target<-ifelse(i<age.bins.N+1,i,
                ifelse(i<age.bins.N+month.N+1,i-age.bins.N,
                    ifelse(i<age.bins.N+month.N+armed.N+1,i-age.bins.N-month.N,
                                i-age.bins.N-month.N-armed.N)))

    for (j in 1:total.levels) {
        cond2.field<-ifelse(j<age.bins.N+1,1,
                    ifelse(j<age.bins.N+month.N+1,2,
                        ifelse(j<age.bins.N+month.N+armed.N+1,3,4)))

        cond2.target<-ifelse(j<age.bins.N+1,j,
                ifelse(j<age.bins.N+month.N+1,j-age.bins.N,
                    ifelse(j<age.bins.N+month.N+armed.N+1,j-age.bins.N-month.N,
                                j-age.bins.N-month.N-armed.N)))

        cond1 <- parse(text=cond.target(cond1.field,cond1.target))
        cond2 <- parse(text=cond.target(cond2.field,cond2.target))

        if(j==i){#Artificially Zero Correlation with Self
            mat.heat[j,i]=0
            next
        }

        mat.heat[j,i]<- nrow(clean.guard.dt[((age>=0) & gender=='Male' & classification =='Gunshot' &
                                            eval(cond1) & eval(cond2))])
    }
}

heatmap(mat.heat+1,labCol=pred.names,labRow = pred.names,
        Rowv = NULL,Colv = NA,scale = 'column',
        main='Male Gunshot Fatalities: Factor Interactions',
        margins=c(9,9),cexRow = 1.25,cexCol = 1.25)
```

Male Gunshot Fatalities: Factor Interactions



Top 50 Most Frequent Condition Pairs

```
top.cond.dt[,.(Occurrences,Condition1,Condition2)]
```

```
##     Occurrences          Condition1     Condition2
##  1:         531          Race:White  Armed:Firearm
##  2:         267          Race:Black  Armed:Firearm
##  3:         164       Armed:Firearm      Age:31-35
##  4:         160 Race:Hispanic/Latino  Armed:Firearm
##  5:         159       Armed:Firearm      Age:26-30
##  6:         145          Race:White      Age:31-35
##  7:         144          Race:White      Age:36-40
##  8:         144          Race:White      Age:36-40
##  9:         130       Armed:Firearm      Age:36-40
## 10:         124          Race:White      Age:26-30
## 11:         122       Armed:Firearm      Age:21-25
## 12:         109          Race:White      Age:46-50
## 13:         102          Race:White      Age:51-55
## 14:         100          Race:Black      Age:26-30
## 15:          99       Armed:Firearm    Month:March
## 16:          98          Race:White      Age:41-45
```

```
## 17:          98           Race:White       Age:41-45
## 18:          96           Race:White     Month:March
## 19:          96           Race:White     Month:March
## 20:          94           Race:Black       Age:21-25
## 21:          93           Race:White      Month:July
## 22:          89        Armed:Firearm      Month:June
## 23:          88        Armed:Firearm   Month:January
## 24:          86        Armed:Firearm Month:September
## 25:          85        Armed:Firearm       Age:41-45
## 26:          85        Armed:Firearm       Age:41-45
## 27:          84           Race:Black       Age:31-35
## 28:          83        Armed:Firearm  Month:February
## 29:          82        Armed:Firearm       Age:51-55
## 30:          82        Armed:Firearm       Age:51-55
## 31:          81           Race:White   Armed:Unknown
## 32:          80           Race:White     Month:April
## 33:          80           Race:White     Month:April
## 34:          79           Race:White       Age:21-25
## 35:          79           Race:White       Age:21-25
## 36:          79           Race:White       Age:21-25
## 37:          79           Race:White       Age:21-25
## 38:          79           Race:White       Age:21-25
## 39:          78        Armed:Firearm     Month:April
## 40:          76           Race:White   Month:October
## 41:          76           Race:White   Month:October
## 42:          75           Race:White       Age:56-60
## 43:          75           Race:White       Age:56-60
## 44:          75           Race:White       Age:56-60
## 45:          74        Armed:Firearm       Age:46-50
## 46:          74        Armed:Firearm       Age:46-50
## 47:          73           Race:White  Month:November
## 48:          69           Race:White      Month:June
## 49:          68 Race:Hispanic/Latino       Age:26-30
## 50:          66           Race:Black        Armed:No
##     Occurrences           Condition1      Condition2
```