Bayes Project

Bianca, Heather, Casey

3/5/2018

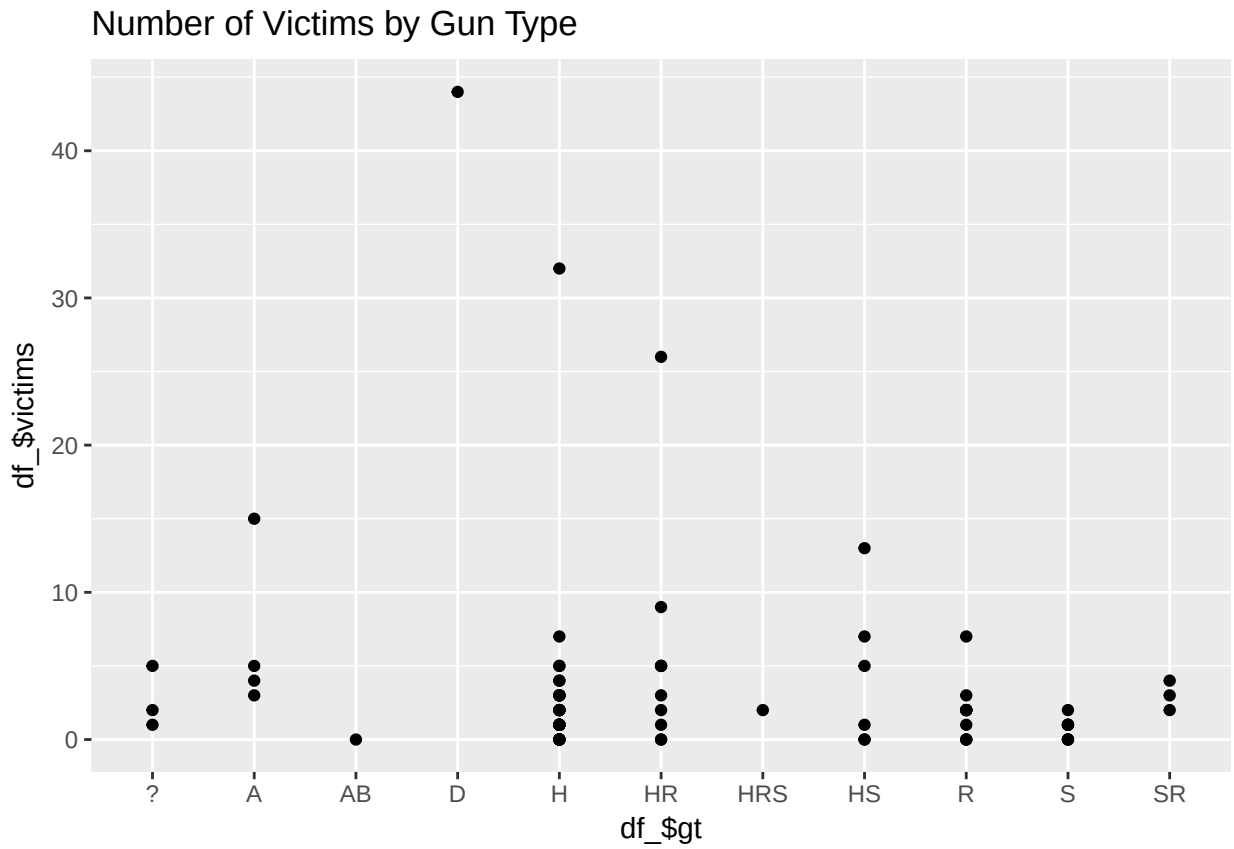
Intro

School shootings have long been unique in America for their frequency and intensity. We have obtained a dataset on all US school shootings since 1766 through UMass GRID in order to begin research on an issue that has had very little scientific study. By examining this data-set in a statistically rigorous way, we hope to gain a better understanding of a tragic phenomenon that does not seem to be in decline.

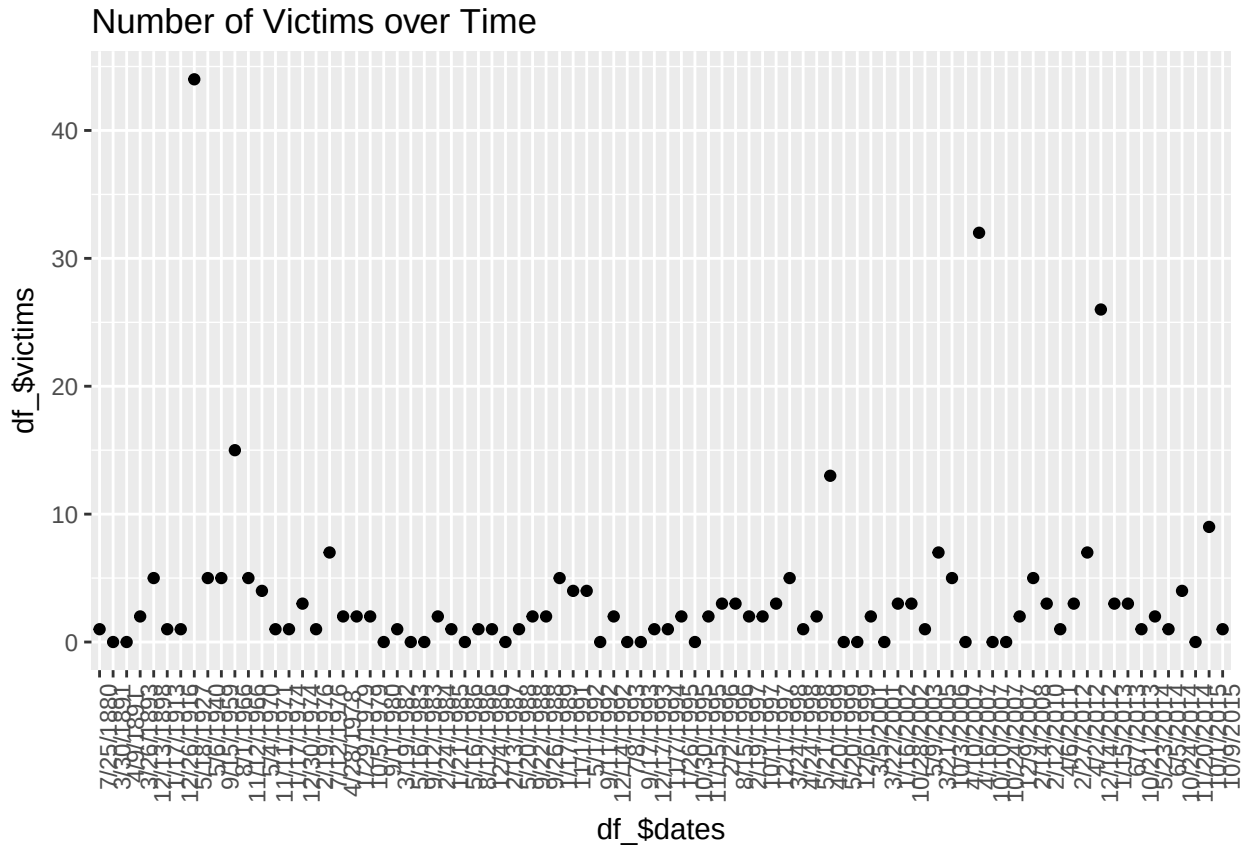
Data

```
dates <- c("7/25/1880","3/30/1891","4/9/1891 ","3/26/1893","12/13/1898","1/17/1913","12/26/1916","5/18/

dates_formatted <- c()

for (i in 1:length(dates)){
  dates_formatted <- c(dates_formatted,as.Date(dates[i],"%m/%d/%Y"))
}

city <- c("Baton Rouge","Liberty","Newburgh","Plain Dealing","Charlestown","Honolulu","Lincoln ","Bath"


county  <- c("East Baton Rouge Parish","Amite","Orange","Bossier Parish","Poca","Honolulu","Casey","Cli


state <- c("LA","MS","NY","LA","WV","HI","KY","MI","CA","TX","TX","AZ","OH","WA","IL","NY","CA","CA","I

victims <- c("1","0","0","2","5","1","1","44","5","5","15","5","4","1","1","3","1","7","2","2","2","0",

r <- c("W","?","W","W","?","L","W","W","W","W","W","W","?","W","W","W","W","W","?","W","A","","W","W","


gt <- c("HS","S","S","?","?","S","H","D","H","HR","A","H","A","R","H","SR","S","R","H","R","H","S","H",

df_ <- data.frame(id = 1:84)

df_$dates <- as.factor(dates)
df_$dates <- factor(df_$dates , levels = dates)

df_$gt <- as.factor(gt)
df_$r <- r
df_$victims <- as.numeric(victims)
df_$state <- state
df_$count <- county
df_$city <- city
library(ggplot2)
```

```
ggplot(df_,aes(x=df_$gt,y = df_$victims)) + geom_point()    + labs(title = "Number of Victims by Gun Type
```




```
ggplot(df_,aes(x=df_$dates,y = df_$victims)) + geom_point() + labs(title = "Number of Victims over Time
```



Potential model: Let

$$y_i = \text{ number of victims in incident}$$

$$r_i = \text{ race of the perpetrator}$$

$$g_i = \text{ gun used by perpetrator}$$

$$s_i = \text{ type of institution}$$

$$\alpha_j[i] = \text{ random intercept accounting for geographic region}$$

$$Y_i \sim Pois(\lambda_i)$$

$$log(\lambda_i) = \alpha_{j[i]} + \beta_1 * r_i + \beta_2 * g_i + \beta_3 * t_i + \beta_4 * s_i$$

```
library(rjags)
```

```
## Loading required package: coda
## Linked to JAGS 4.3.0
## Loaded modules: basemod,bugs
```

```
library(R2jags)
```

```
##
## Attaching package: 'R2jags'
```

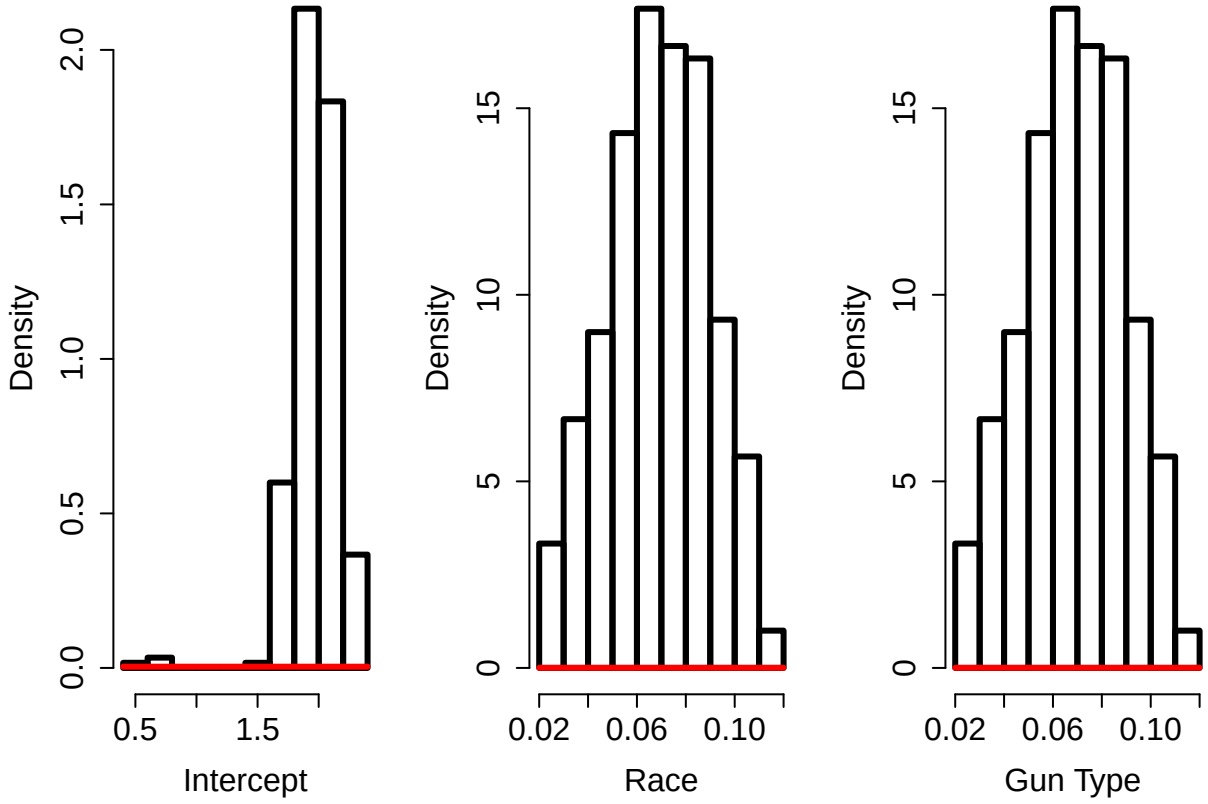
```
## The following object is masked from 'package:coda':
##
##     traceplot
```

```
model <- "
model {
for( i in 1:n) {
log(lambda.i[i]) <- (
# different notation; using betas for main effects and b's for varying effects
  b0
  + b1*r[i] + b2*gt[i]
  )
  y.i[i] ~ dpois(lambda.i[i])
}

b0  ~ dnorm(0.0,1.0E-4)
b1 ~ dnorm(0.0,1.0E-4)
b2 ~ dnorm(0.0,1.0E-4)

} # end model
"
jags.data <- list(y.i = df_$victims,
                  r = as.factor(df_$r), gt =as.factor(df_$gt),  n=length(df_$victims))

parnames <- c( "b1", "b0","b2")
mod <- jags(data = jags.data,
                   parameters.to.save=parnames,
                   n.chains = 3, n.burnin = 1500, n.iter =1500 + 1000, n.thin = 10, model.file = textC
```

```
## module glm loaded
```

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 84
##    Unobserved stochastic nodes: 3
##    Total graph size: 338
##
## Initializing model
```

```
# some priors and posteriors to check that priors weren't informative
mcmc.array <- mod$BUGSoutput$sims.array
par(lwd = 3, cex.axis = 1.5, cex.lab = 1.5, cex.main = 1.5,mar = c(5,5,1,1), mfrow = c(1,3))
hist(c(mcmc.array[,,"b0"]), freq = F, main = "", xlab ="Intercept")
curve(dnorm(x,0,sqrt(1/10^(-4))), add = T, col = 2)

hist(c(mcmc.array[,,"b1"]), freq = F, main = "", xlab ="Race")
curve(dnorm(x,0,sqrt(1/10^(-4))), add = T, col = 2)

hist(c(mcmc.array[,,"b1"]), freq = F, main = "", xlab ="Gun Type ")
curve(dnorm(x,0,sqrt(1/10^(-4))), add = T, col = 2)
```



We will also look at interactions and possible transformations of covariates as needed.

Questions:
- Would a time series component work for this model?
- Should institution type be a hierarchical structure?