Progress Report - Are school shootings becoming more frequent?

Bianca Doone

April 3, 2018

Questions:

Are the frequency of shootings really increasing?

Whare are the projections for shootings in the next 5,10,50 years? How many shootings can we expect this year? Are we above average right now?

The data I'm working with:

- 27 years of data, plus up to March 19th, 2018
- Incidences of school shootings
- Created a variable to show the amount of days between shootings
- Combined data into year format (number of shootings per year, maximum number of days without a shooting, proportion of shootings per year)

Note: I have done a lot of data cleaning/organizing/restructuring done from the original dataset.

Goal of My Analysis:

The plan is to model the number shootings per year with a Poisson distribution, and a prior for lambda that is an AR(1) time series. Report results in proportions of shooting per year.

Then I plan to forcast the year 2018 and compare it to the actual data we have now.

Current Problems:

- I'm not sure if my proposed model is suitable for the type of data I'm working with (not particularly sure how to model rare-binary outcomes in a time-series)
- I can't get my JAGS code to run, the code is shown below
- I still haven't learned how to do in and out of sample validation to compare my forcast with the actual data
- Is my proposed plan enough for my segment?

My Analysis

Summary Statistics (TOTAL)

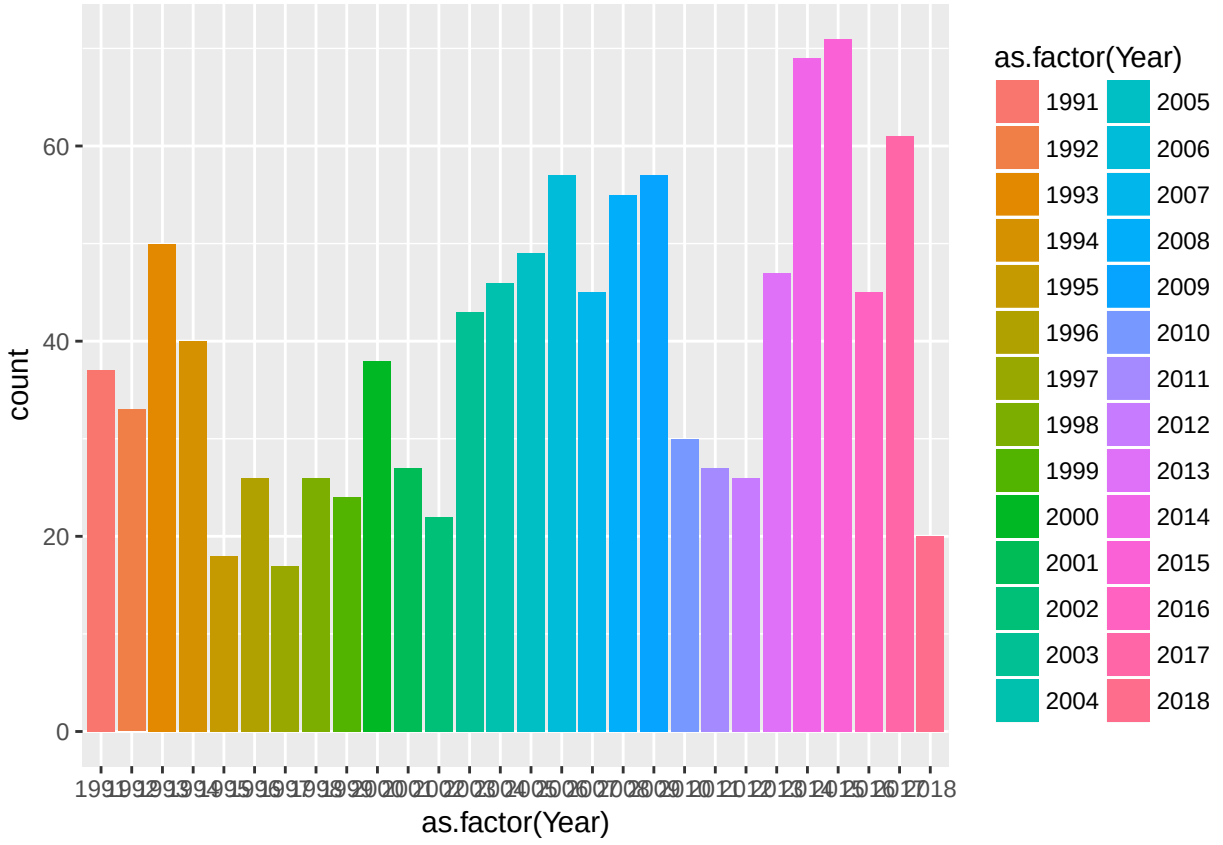
```
table(ss_final$Shooting)
```

```
##
##    0    1
## 8924 1016
```

```
#add table year and number shootings/proportion
```

Incidence of Shooting Over Time

```
#basic histogram - number of shootings per year (with proportions table)
ggplot(data=shootings2, aes(x = as.factor(Year), fill = as.factor(Year))) +
  geom_bar()
```

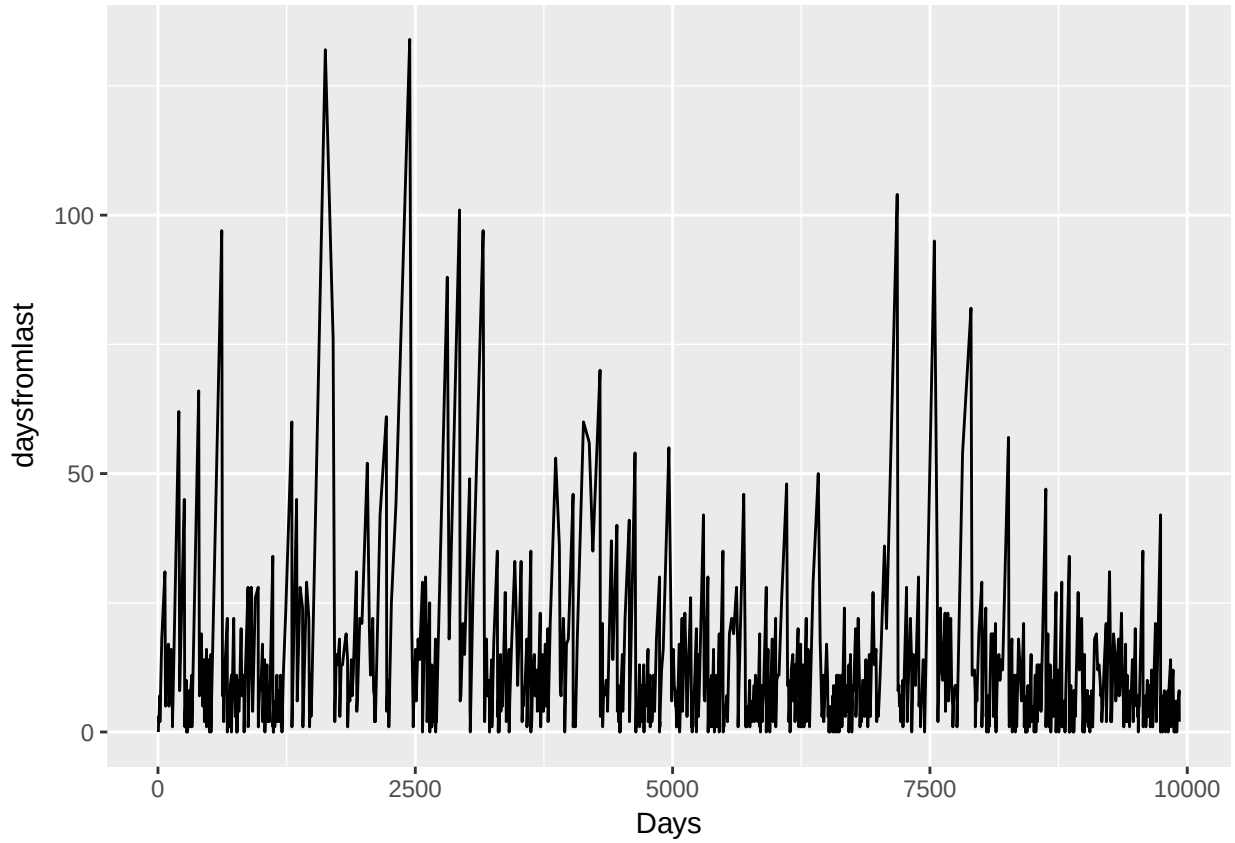



```
print(year_df$Prop.Shootings.Year)
```

```
##  [1] 0.09589041 0.07923497 0.12328767 0.10136986 0.04931507 0.07103825
##  [7] 0.04657534 0.05753425 0.06027397 0.09289617 0.07123288 0.06027397
## [13] 0.10684932 0.11748634 0.11780822 0.15342466 0.10958904 0.13934426
```

```
## [19] 0.13972603 0.08219178 0.06849315 0.07103825 0.11232877 0.16438356
## [25] 0.16986301 0.12295082 0.12328767 0.20512821
```

```
#basic graph - days from last shooting against shooting (with table maximums)
ggplot(data = shootings2, aes(x = Days, y = daysfromlast)) + geom_line()
```




```
print(year_df$Max.difference)
```

```
##  [1]  62  97  28  60 132  52 134  88 101  35  53  70  54  55  42  46  48
## [18]  50  24 104  95  82  57  47  34  31  42   8
```

My Model

Shooting time-series model (by year):

$$y_t \sim \text{Poisson}(\lambda_t)$$

$$\lambda_t \sim \text{AR}(1) = \rho\lambda_{t-1} + \epsilon_t$$

where y_t is the number of shootings in year t.

I couldn't get this to run :(

```
GetAR <- function(nyears, # length of series
                  rho, sigma, # AR parameters
                  eps0.t = NULL, # innovations (optional)
                  ystart = NULL # starting value y1 (optional)
                  ){
  if (is.null(eps0.t)){
    set.seed(123)
    eps0.t <- rnorm(nyears, 0, 1)
  }
  y.t <- rep(NA, nyears)
  if (is.null(ystart)){
    y.t[1] <- sigma/sqrt(1-rho^2)*eps0.t[1]
  } else {
    y.t[1] <- ystart
  }
  for (t in 2:nyears){
    y.t[t] <- rho*y.t[t-1] + sigma*eps0.t[t]
  }
  return(y.t)
}

sigma <- 1
rho <- 0.5
nyears <- year_df$Year
jags.data <- list(y.t = year_df$Shootings, nyears=year_df$Year)
parnames <- c("sigma", "rho")
set.seed(1)

model <- "model{
y.t[1] ~ dpois(lambda.t[1])
lambda.t[1] ~ dnorm(0,tau.lambda)
tau.lambda <- (1-rho^2)/sigma^2
for (t in 2:nyears){
 y.t[t] ~ dpois(lambda.t[t])
 lambda.t[t] <- rho*y.t[t-1]
}

sigma ~ dunif(0,15)
rho ~ dunif(-1,1)
}"
P = 1

mod <-jags(data = jags.data,
                    parameters.to.save= c("y.t", parnames),
           model.file = textConnection(model))
max(mod$BUGSoutput$summary[, c("Rhat")])
min(mod$BUGSoutput$summary[, c("n.eff")])
round(mod$BUGSoutput$summary[c("sigma",  "rho", "lambda.t"),],2)

ypred.sp <- mod$BUGSoutput$sims.list[["y.p"]]
ypred.qp <- apply(ypred.sp, 2, quantile, c(0.025, 0.5, 0.975))
```

Comparing Pre- and Post-2018 (Using either in or out of sample validation)