

Modeling K-12 school shootings as a clustered point process and an areal process, from 1990-2019

J Steven Raquel

Department of Statistics, University of California, Irvine

March 13, 2022

Abstract

Keywords:

1 Background

The prevalence of gun violence in schools in the United States has been referred to both as an “epidemic” and a public health crisis, and one that has steadily increased over the past several decades (see Figure 1). The debate on how to curb these tragedies is a solidly partisan issue, with calls for more and stronger gun control laws, as well as more mental health care, sometimes as an alternative to gun control. Concealed carry laws are argued both in favor of, and against and are labeled as both an enabler and a potential deterrent to mass shootings. Apart from the trauma that such an event can bring to a community, there is also resonant fear that such incidents inspire “copycat” events on a local and a larger scale. This spatial analysis attempts to model the incidence of these shootings as a Poisson point process, in order to ascertain whether the locations and events occur with complete spatial randomness. We also model the data in an areal form on a per-state basis so as to compare it against the level of gun control and mental health care, also on a state level.

1.1 Data

The school shooting data was sourced directly from the “K-12 School Shooting Database” made available by the Center for Homeland Defense and Security (CHDS). The information that comprises the dataset was determined by a specific process which entailed asking what exactly comprises a school shooting. Most shootings that were premeditated and resulted in mass casualties, such as those that occurred at Columbine in 1999, Sandy Hook in 2012, and Stoneman Douglas in 2018, are considered “indiscriminate shootings”. Although the original database contains shootings ruled accidental (from misuse of a firearm) as well as incidences of gang-related gun violence on school grounds, among other incidents, we did not opt to consider this data as relevant to this study in particular.

It was a deliberate decision on our part to filter out these incidents as others like it so as to hone in on the more specific and culturally recognized definition of a school shooting, which is an attack that takes place on school grounds, in order to target students, faculty, or passersby with the intent to cause terror and/or inflict harm on specific individuals. Targeted events related to domestic situations (e.g. anger over a break-up), or the escalation

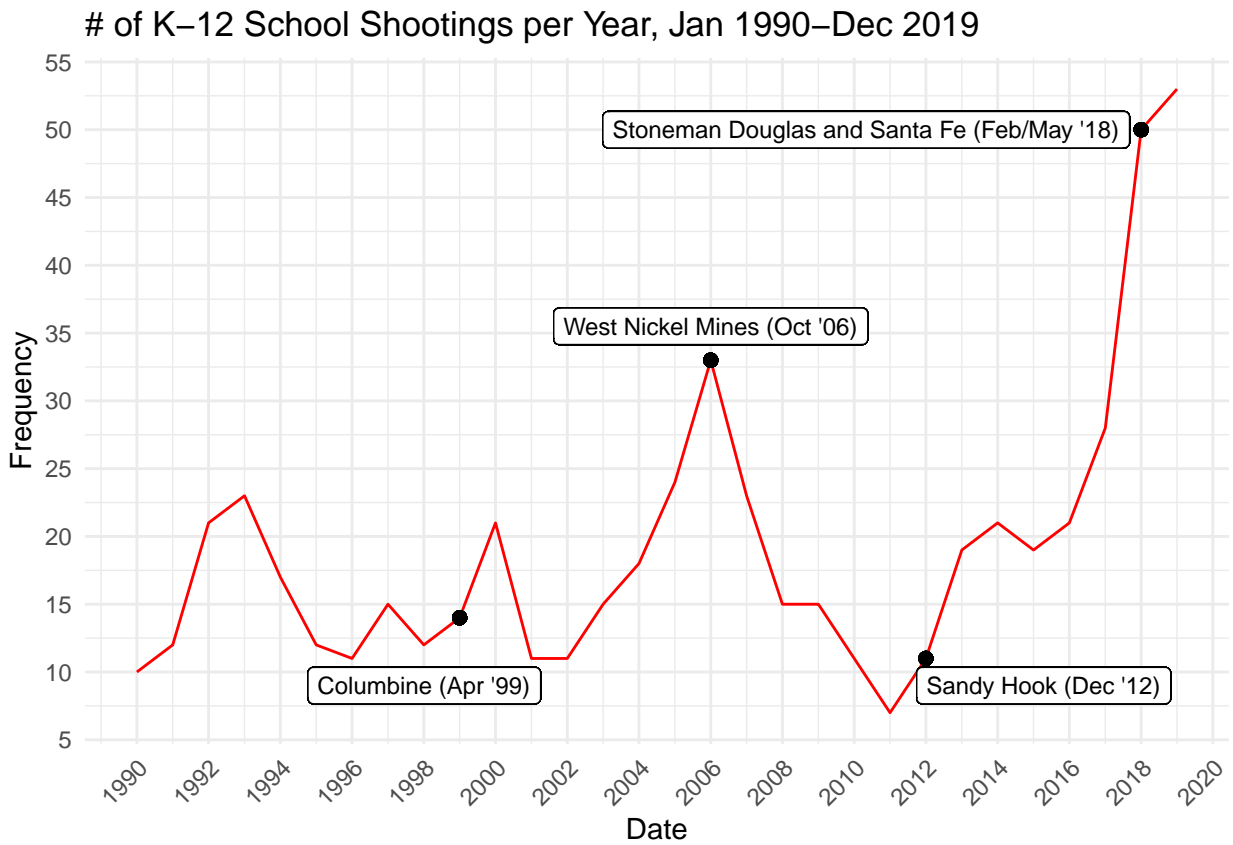


Figure 1: Time series plot of number of K-12 school shootings per year, from Jan 1990 to Dec 2019.

of disputes (e.g. fistfights in which one person pulls out a firearm) were also ruled school shootings.

The database originally goes as far back as 1970 all the way to present day (March 2022 at the time of this writing), but it was of particular interest to look at the data after 1990 since this consolidates the modern era in which access to firearms and the incidence of gun violence in schools is more normalized. It was also decided to leave out data past 2019 as the COVID-19 pandemic which hit the US in 2020 is a huge confounding variable as it caused many schools to institute remote learning which decreased the incidence of shootings on school campuses.

The data originally only contained the name of the school, as well as the city and state where the events took place, but we were able to use the Google Maps API to source the specific latitude/longitude coordinates of these events.

In addition to this school shooting data, we also have gun control data consisting of several columns of binary categorical data as to whether or not a certain law (e.g. a ban on concealed carry, background checks) is instituted in a given state for the years between 1991 and 2020 from StateFirearmLaws.org, as well as mental healthcare data from 2019 courtesy of Mental Health America (MHA), which consists of a ranking of all 50 states and the District of Columbia on their ratio of need for and accessibility to mental healthcare. More information on these resources can be found on their respective websites (see References section).

2 Methods

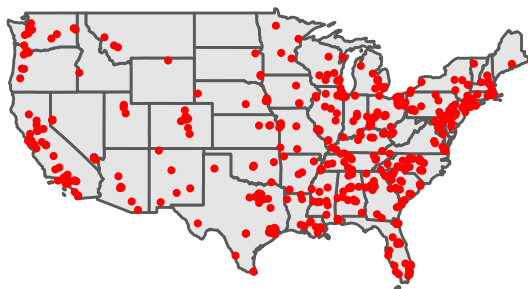
2.1 Exploratory Data Analysis

As a preliminary analysis, we chose to plot this data on both a point and at an areal level, for a number of reasons. For one, modeling the data as a point pattern allows us to test the null hypothesis for complete spatial randomness against the alternative hypothesis of clustering. It allows us to hone in on exactly where within a given state might be considered a “hotspot” for a shooting. Conversely, the areal data allows us to compare the number of incidents in a given state and look at it with respect to the level of gun control and mental

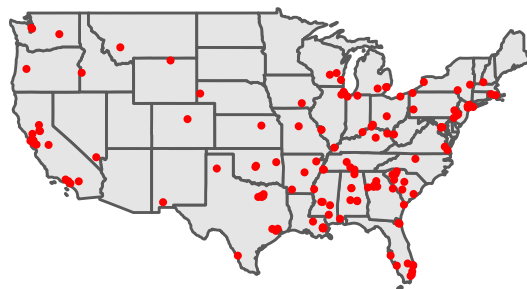
healthcare available there as well.

Shootings in K–12 schools in the US

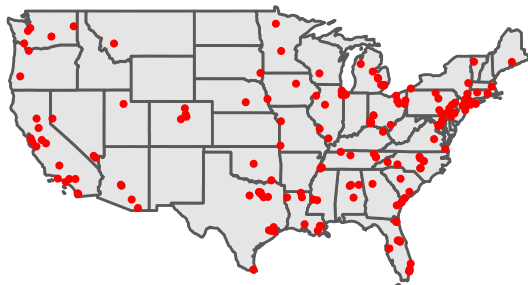
1990–2019, Total: 573



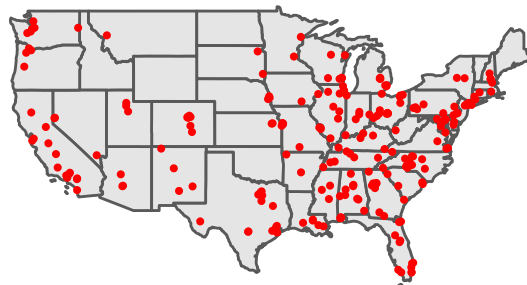
1990–99, Total: 147



2000–09, Total: 186

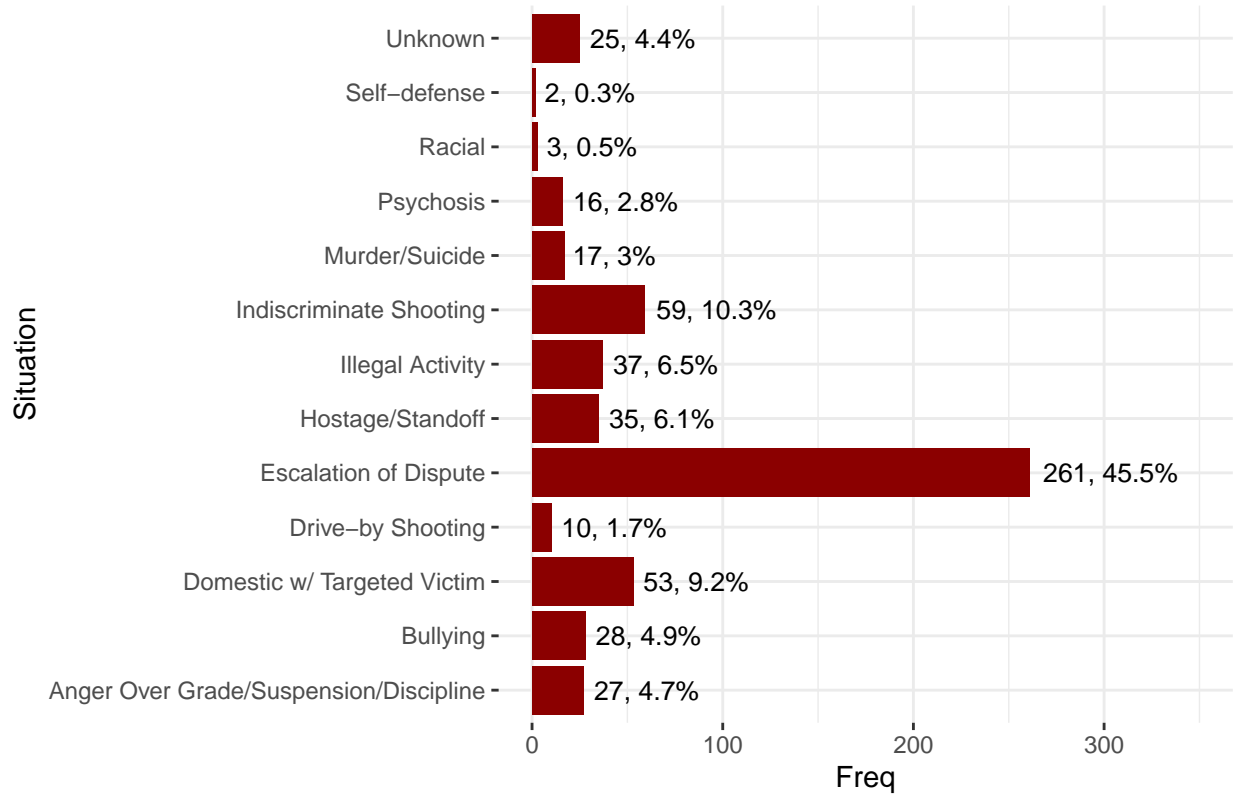


2010–19, Total: 240

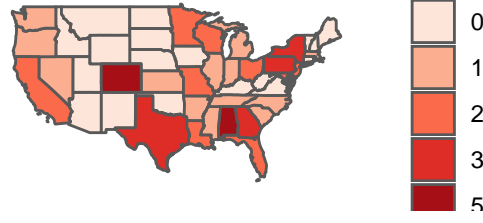


As we can see from the

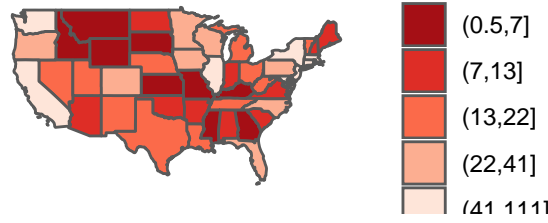
K–12 School Shootings, by Situation, 1990–2019



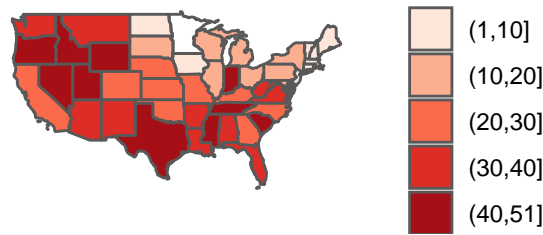
Number of School Shootings per State, 2019



Number of Gun Control Laws per State, 2019



Ranking of Mental Healthcare by State, 2019



2.2 Point Pattern Analysis

As with any spatial point pattern analysis, we are concerned with the following three questions, 1) whether the points are located at random, 2) whether they are clustered, and 3) whether they are placed regularly. The hypothesis of *complete spatial randomness* (CSR) asserts the following:

- The number of events in any region S with area $|S|$ follows a Poisson distribution with mean $\lambda|S|$, where λ is the intensity, i.e. λ does not change over S
- Given n events in S , the points s_i are independently located according to the uniform distribution on S , i.e. there is no interaction amongst events.

The intensity function $\lambda(s)$, also known as the first-order property of the spatial point process, is defined as

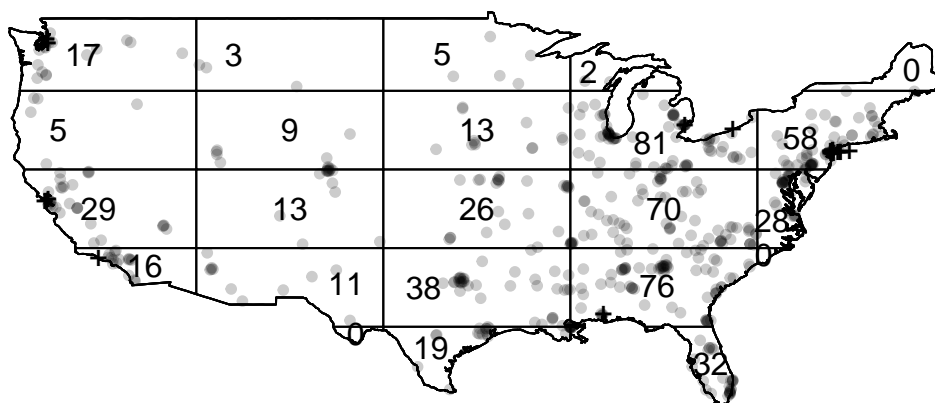
$$\lambda(s) = \lim_{|\Delta s| \rightarrow 0} \frac{E[N(\Delta s)]}{|\Delta s|}$$

Firstly we want to ascertain whether the incidences of school shootings are indeed a Poisson process, and if so, determine whether or not the process is *homogeneous* (where the intensity function $\lambda(s)$ assumes a constant λ) or *inhomogeneous*. Contextually, this means we are interested in ascertaining whether the spatial pattern of these shootings is random or not, i.e. are they more likely to take place in certain places, or around each other? The intensity function is the expected number of events per unit area.

2.3 Quadrats

While we could estimate the intensity function across the entire area, what we are interested in this particular analysis is to how the intensity varies across different regions contained therein. We do this by splitting up the area into what are referred to as *quadrats*.

Poisson Point Pattern, School Shootings 1990–2019



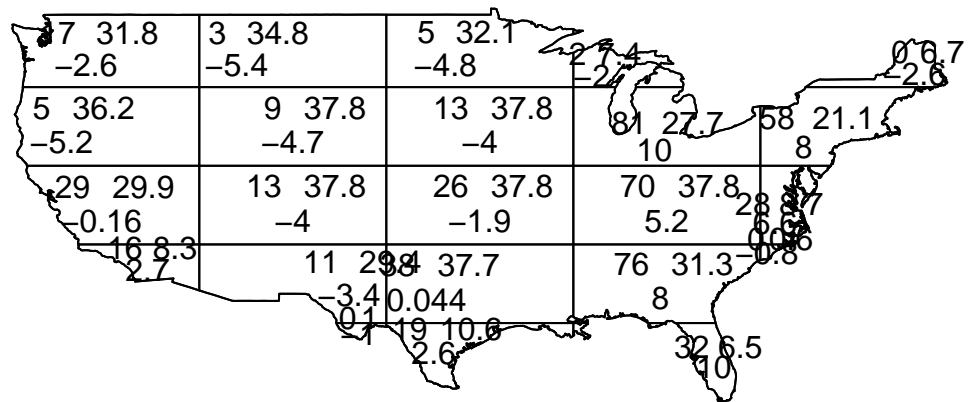
##

Conditional Monte Carlo test of CSR using quadrat counts

Test statistic: Pearson X2 statistic

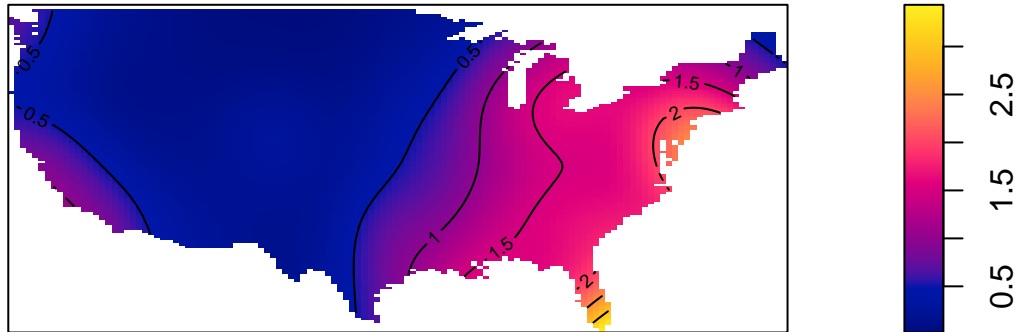

```
##
## data:  ppp_9019
## X2 = 582.1, p-value = 2e-04
## alternative hypothesis: clustered
##
## Quadrats: 23 tiles (irregular windows)
```

Results of Quadrat Test, per Quadrat

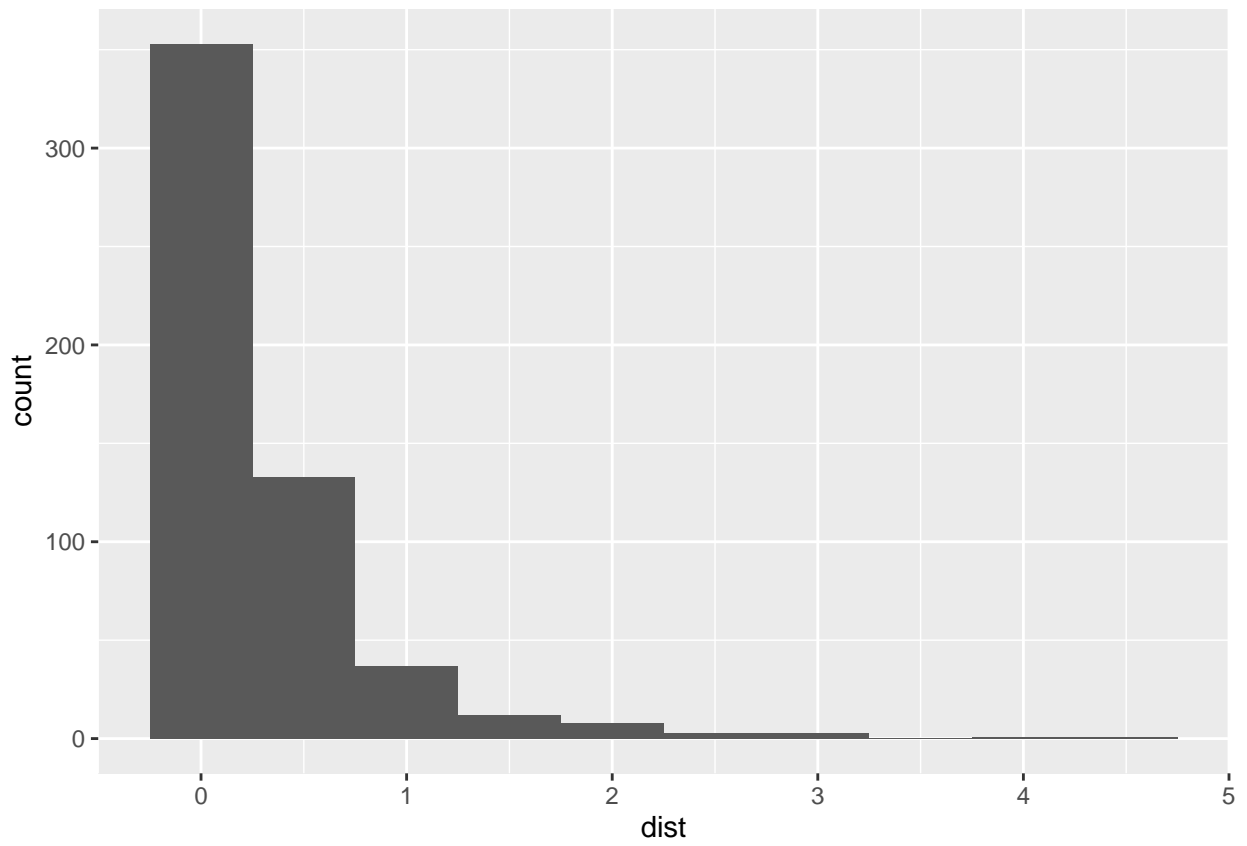


```
## Kernel Density
```

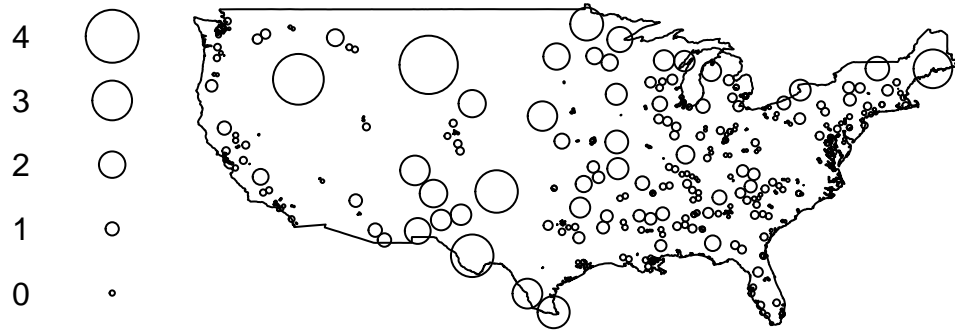
Density of School Shootings, 1990–2019



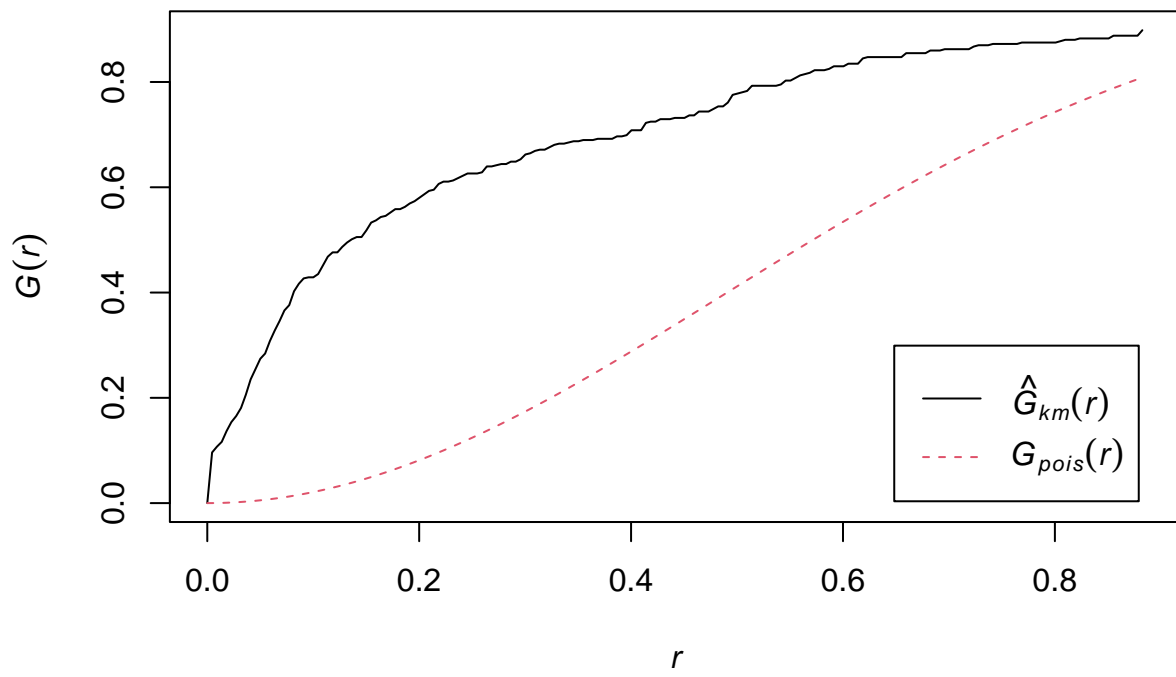
2.4 Nearest Neighbor



Stienen diagram, 1990–2019 shootings

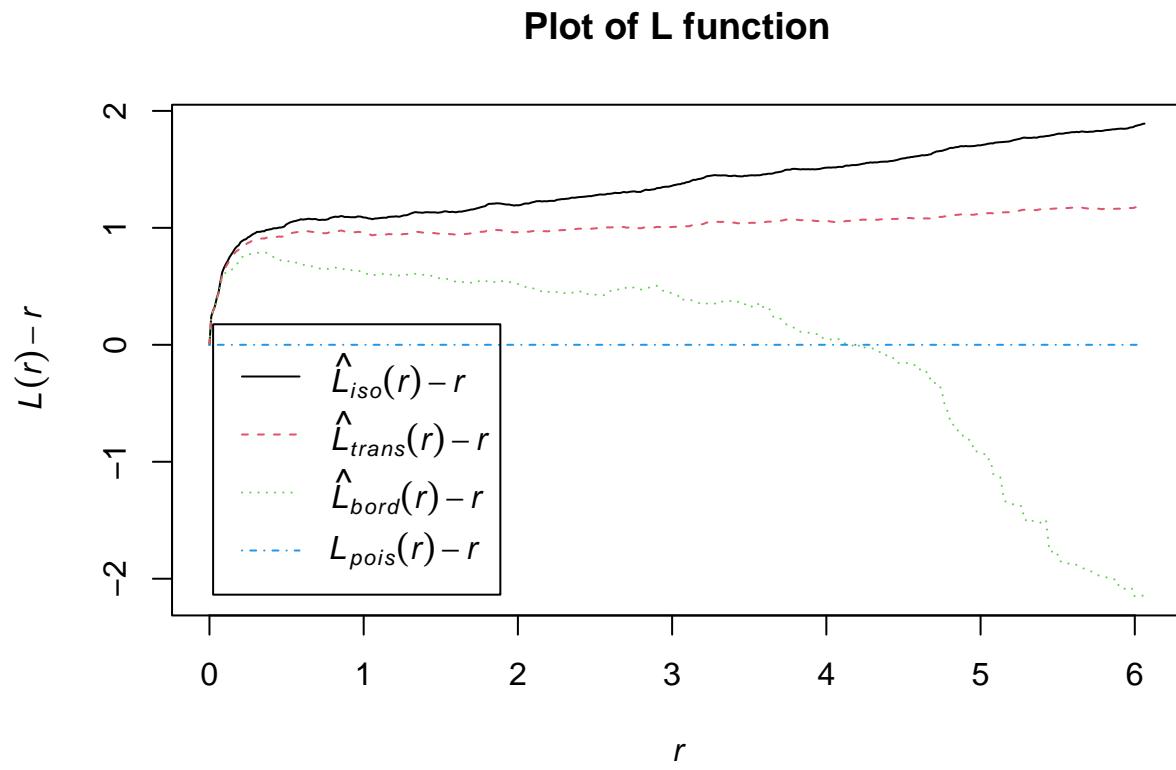


G function



3 Discussion

4 References



5 References