PROBABILISTIC METHODS IN ENGINEERING
(VE-401)

TERM PROJECT REPORT 2

2017 Spring Team 8

Name:		ID:	
Name:	He Kaijun	ID:	515370910211
	Kang Yiyan		5143709052
	Sheng Beiyi		515370910110
	Yang Zixuan		515370910217
	Zhou Yimeng		515370910085

Instructor:	Horst Hohberger
Teaching Assistant:	Wang Zesen
	Zhu Weiyu

Date: April 24, 2017

Contents

Abstract

In this project, we analyze the occurrence of fatal police shootings in the US from 2015 to 2017 based on the data of the Washington Post [1] with the analogous manner that used in the article *London murders: a predictable pattern?* by David Spiegelhalter and Arthur Barnett [2]. We summary the data source, make some conclusions after sorting the data due to different days, weekdays, months and reasons of shooting, and predict the number of shooting per day and the gap width between shootings.

1 Objectives

This project is aimed at applying our basic knowledge of probabilistic methods in engineering including probability and statistics to several interesting issues beyond class assignments. It requires our team efforts on each question, involving comprehension, data collection, discussion and, more importantly, the balance between individual thoughts and team collaboration.

2 The Data Source

In our project, we will basically use the data acquired from the *Database of Fatal Police Shootings* of the Washington Post[1]. It shows accurate statistics about fatal shootings in the United States committed by police officers from January 1st, 2015 to present.

It is worthy mentioning that the improvements the Post did in 2015 to perfect the statistics and details as much as they can has made the situation more clearly and useful to some extent. In detail, it began collecting several details, such as race, whether the person was armed and whether the person was fleeing. In addition, the Officers' names who were involved in the shooting are also included in the database in 2016. These changes in the completeness of the data shows the increasing reliability and usefulness for people hoping to conduct research based on those data.

Although the database is updated every day if there is new data emerging, it is still encountering doubts from the FBI. Even in last year, the FBI stated that plans were carried out to "overhaul" the tracking methods[1].

Nevertheless, the term "fatal police shooting" means the shooting by police which caused death, which should be a useful database for the police and even the government.

3 Daily Police Shooting Data In 2015 & 2016

Since the abundant single data is not clear enough for statistics, we should combine some of the major information to see the trend of the police shooting data. In this part, we will first focus on the daily quantity in the two entire years, both 2015 and 2016.

3.1 Figure Generated By Original Data

We use the original data in the database to generate the figure using Mathematica. The data is a column of dates consisting of repeating and different ones according to each shooting circumstance. The figure is shown as Figure [1]

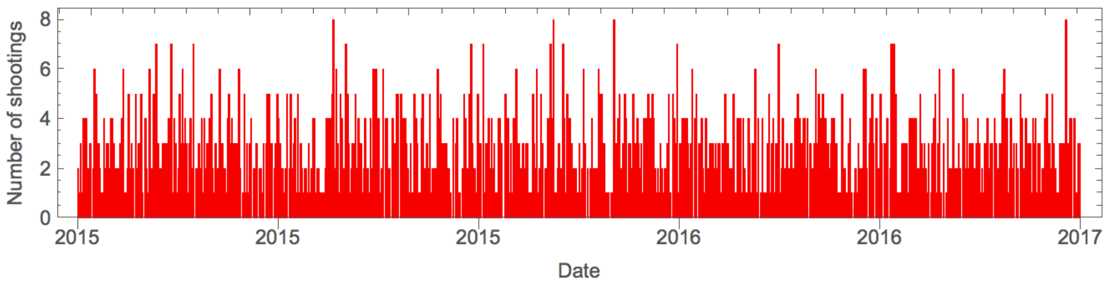


Figure 1: Daily Fatal Police Shooting Data from original data by Mathematica

3.2 Figure Verified By Manual Data Collecting

We did take some time collecting the statistical data, and have created two relative tables for the two years, for the sake of clarity. Here we have to mention that since 2016 is a leap year while 2015 is not, to maintain the uniformity of the data, we set February 29^{th}, 2015 as invalid date, and at the same time, set the count as zero, which is labeled red in the Table [1].

Table 1: Daily Fatal Police Shooting Data in 2015

day/month	No.	day/month	No.	day/month	No.	day/month	No.	day/month	No.
01/01	0	03/15	1	05/27	5	08/08	4	10/20	5
01/02	2	03/16	2	05/28	4	08/09	4	10/21	3
01/03	1	03/17	5	05/29	3	08/10	1	10/22	3
01/04	3	03/18	3	05/30	2	08/11	2	10/23	0
01/05	1	03/19	6	05/31	2	08/12	6	10/24	7
01/06	4	03/20	3	06/01	1	08/13	0	10/25	1
01/07	4	03/21	4	06/02	2	08/14	4	10/26	3
01/08	4	03/22	3	06/03	5	08/15	4	10/27	1
01/09	2	03/23	1	06/04	0	08/16	2	10/28	4
01/10	0	03/24	3	06/05	2	08/17	2	10/29	2
01/11	3	03/25	4	06/06	5	08/18	0	10/30	4
01/12	0	03/26	2	06/07	2	08/19	4	10/31	2
01/13	2	03/27	7	06/08	3	08/20	3	11/01	3
01/14	6	03/28	0	06/09	4	08/21	3	11/02	1
01/15	5	03/29	2	06/10	1	08/22	5	11/03	1
01/16	2	03/30	2	06/11	5	08/23	5	11/04	4

day/month	No.	day/month	No.	day/month	No.	day/month	No.	day/month	No.
01/17	3	03/31	3	06/12	1	08/24	0	11/05	4
01/18	2	04/01	2	06/13	3	08/25	5	11/06	3
01/19	1	04/02	4	06/14	2	08/26	3	11/07	0
01/20	1	04/03	1	06/15	1	08/27	4	11/08	0
01/21	4	04/04	4	06/16	2	08/28	4	11/09	5
01/22	2	04/05	2	06/17	2	08/29	3	11/10	4
01/23	3	04/06	3	06/18	1	08/30	2	11/11	4
01/24	3	04/07	1	06/19	3	08/31	1	11/12	1
01/25	2	04/08	4	06/20	1	09/01	3	11/13	3
01/26	4	04/09	5	06/21	2	09/02	1	11/14	1
01/27	4	04/10	2	06/22	4	09/03	1	11/15	4
01/28	3	04/11	1	06/23	4	09/04	5	11/16	2
01/29	2	04/12	3	06/24	1	09/05	4	11/17	6
01/30	1	04/13	1	06/25	2	09/06	3	11/18	2
01/31	2	04/14	2	06/26	2	09/07	1	11/19	3
02/01	0	04/15	6	06/27	1	09/08	1	11/20	2
02/02	3	04/16	1	06/28	1	09/09	4	11/21	1
02/03	4	04/17	3	06/29	1	09/10	5	11/22	3
02/04	6	04/18	2	06/30	1	09/11	2	11/23	2
02/05	1	04/19	2	07/01	2	09/12	2	11/24	3
02/06	1	04/20	1	07/02	4	09/13	3	11/25	3
02/07	2	04/21	4	07/03	4	09/14	2	11/26	1
02/08	5	04/22	5	07/04	4	09/15	5	11/27	1
02/09	2	04/23	3	07/05	4	09/16	1	11/28	1
02/10	3	04/24	4	07/06	5	09/17	2	11/29	5
02/11	2	04/25	3	07/07	8	09/18	2	11/30	3
02/12	0	04/26	3	07/08	2	09/19	2	12/01	2
02/13	5	04/27	1	07/09	6	09/20	2	12/02	6
02/14	2	04/28	3	07/10	3	09/21	6	12/03	1
02/15	3	04/29	6	07/11	1	09/22	4	12/04	2
02/16	3	04/30	2	07/12	5	09/23	5	12/05	5
02/17	5	05/01	0	07/13	4	09/24	3	12/06	3
02/18	1	05/02	1	07/14	2	09/25	3	12/07	0
02/19	0	05/03	5	07/15	1	09/26	2	12/08	2
02/20	4	05/04	1	07/16	7	09/27	3	12/09	2
02/21	2	05/05	3	07/17	5	09/28	2	12/10	4
02/22	1	05/06	1	07/18	3	09/29	2	12/11	1
02/23	6	05/07	3	07/19	1	09/30	1	12/12	7
02/24	1	05/08	4	07/20	2	10/01	0	12/13	4
02/25	2	05/09	0	07/21	3	10/02	4	12/14	8
02/26	5	05/10	2	07/22	3	10/03	0	12/15	1
02/27	1	05/11	2	07/23	5	10/04	3	12/16	1
02/28	7	05/12	3	07/24	2	10/05	4	12/17	2
02/29	0	05/13	0	07/25	4	10/06	1	12/18	3
03/01	3	05/14	2	07/26	2	10/07	1	12/19	2
03/02	1	05/15	2	07/27	3	10/08	0	12/20	2
03/03	2	05/16	1	07/28	1	10/09	2	12/21	7

day/month	No.	day/month	No.	day/month	No.	day/month	No.	day/month	No.
03/04	2	05/17	3	07/29	2	10/10	5	12/22	4
03/05	3	05/18	0	07/30	4	10/11	4	12/23	1
03/06	3	05/19	3	07/31	2	10/12	2	12/24	5
03/07	1	05/20	5	08/01	2	10/13	0	12/25	1
03/08	3	05/21	5	08/02	1	10/14	5	12/26	4
03/09	4	05/22	1	08/03	4	10/15	7	12/27	3
03/10	4	05/23	3	08/04	2	10/16	3	12/28	1
03/11	7	05/24	0	08/05	6	10/17	2	12/29	2
03/12	1	05/25	2	08/06	2	10/18	2	12/30	2
03/13	4	05/26	3	08/07	6	10/19	1	12/31	1
03/14	3								

Table 2: Daily Fatal Police Shooting Data in 2016

day/month	No.	day/month	No.	day/month	No.	day/month	No.	day/month	No.
01/01	1	03/15	3	05/27	1	08/08	5	10/20	1
01/02	3	03/16	3	05/28	2	08/09	3	10/21	2
01/03	1	03/17	4	05/29	4	08/10	0	10/22	1
01/04	2	03/18	2	05/30	2	08/11	3	10/23	3
01/05	6	03/19	4	05/31	1	08/12	2	10/24	1
01/06	3	03/20	4	06/01	3	08/13	4	10/25	4
01/07	0	03/21	3	06/02	2	08/14	0	10/26	2
01/08	2	03/22	2	06/03	2	08/15	2	10/27	3
01/09	0	03/23	3	06/04	3	08/16	7	10/28	3
01/10	2	03/24	6	06/05	1	08/17	1	10/29	0
01/11	4	03/25	0	06/06	2	08/18	7	10/30	4
01/12	2	03/26	4	06/07	4	08/19	5	10/31	3
01/13	2	03/27	5	06/08	1	08/20	1	11/01	1
01/14	2	03/28	0	06/09	5	08/21	1	11/02	2
01/15	2	03/29	3	06/10	4	08/22	1	11/03	3
01/16	6	03/30	3	06/11	2	08/23	1	11/04	2
01/17	5	03/31	4	06/12	3	08/24	3	11/05	5
01/18	5	04/01	2	06/13	4	08/25	3	11/06	6
01/19	3	04/02	1	06/14	4	08/26	1	11/07	1
01/20	3	04/03	4	06/15	2	08/27	3	11/08	4
01/21	1	04/04	1	06/16	4	08/28	1	11/09	1
01/22	0	04/05	2	06/17	0	08/29	4	11/10	2
01/23	1	04/06	3	06/18	3	08/30	4	11/11	4
01/24	0	04/07	2	06/19	2	08/31	2	11/12	2
01/25	1	04/08	3	06/20	2	09/01	4	11/13	3
01/26	1	04/09	2	06/21	3	09/02	4	11/14	0
01/27	8	04/10	3	06/22	6	09/03	4	11/15	3
01/28	3	04/11	2	06/23	3	09/04	2	11/16	2
01/29	3	04/12	3	06/24	5	09/05	2	11/17	0
01/30	4	04/13	3	06/25	4	09/06	5	11/18	5
01/31	5	04/14	1	06/26	4	09/07	3	11/19	4
02/01	2	04/15	3	06/27	5	09/08	2	11/20	2

day/month	No.	day/month	No.	day/month	No.	day/month	No.	day/month	No.
02/02	1	04/16	1	06/28	2	09/09	3	11/21	2
02/03	3	04/17	5	06/29	4	09/10	2	11/22	4
02/04	5	04/18	2	06/30	3	09/11	1	11/23	3
02/05	4	04/19	3	07/01	3	09/12	3	11/24	1
02/06	1	04/20	2	07/02	2	09/13	0	11/25	3
02/07	3	04/21	1	07/03	3	09/14	1	11/26	1
02/08	3	04/22	2	07/04	4	09/15	3	11/27	4
02/09	2	04/23	5	07/05	4	09/16	4	11/28	3
02/10	5	04/24	0	07/06	2	09/17	5	11/29	4
02/11	3	04/25	1	07/07	3	09/18	1	11/30	1
02/12	3	04/26	4	07/08	1	09/19	3	12/01	3
02/13	4	04/27	3	07/09	1	09/20	6	12/02	3
02/14	3	04/28	4	07/10	1	09/21	1	12/03	1
02/15	2	04/29	1	07/11	5	09/22	0	12/04	2
02/16	3	04/30	4	07/12	1	09/23	3	12/05	0
02/17	1	05/01	2	07/13	2	09/24	0	12/06	4
02/18	4	05/02	1	07/14	1	09/25	1	12/07	5
02/19	2	05/03	1	07/15	1	09/26	4	12/08	1
02/20	4	05/04	4	07/16	3	09/27	2	12/09	3
02/21	5	05/05	3	07/17	4	09/28	2	12/10	2
02/22	4	05/06	1	07/18	1	09/29	1	12/11	5
02/23	4	05/07	3	07/19	0	09/30	6	12/12	3
02/24	5	05/08	1	07/20	1	10/01	2	12/13	2
02/25	4	05/09	6	07/21	2	10/02	2	12/14	0
02/26	2	05/10	1	07/22	1	10/03	2	12/15	1
02/27	1	05/11	4	07/23	3	10/04	3	12/16	1
02/28	2	05/12	0	07/24	2	10/05	2	12/17	3
02/29	1	05/13	1	07/25	2	10/06	0	12/18	3
03/01	3	05/14	4	07/26	2	10/07	5	12/19	1
03/02	2	05/15	1	07/27	6	10/08	4	12/20	3
03/03	1	05/16	1	07/28	6	10/09	2	12/21	8
03/04	2	05/17	1	07/29	2	10/10	4	12/22	1
03/05	2	05/18	3	07/30	0	10/11	2	12/23	3
03/06	3	05/19	5	07/31	3	10/12	3	12/24	4
03/07	5	05/20	1	08/01	4	10/13	3	12/25	4
03/08	1	05/21	3	08/02	5	10/14	3	12/26	0
03/09	0	05/22	4	08/03	3	10/15	1	12/27	4
03/10	5	05/23	1	08/04	0	10/16	2	12/28	1
03/11	4	05/24	3	08/05	4	10/17	3	12/29	1
03/12	3	05/25	2	08/06	0	10/18	4	12/30	3
03/13	7	05/26	7	08/07	2	10/19	3	12/31	3
03/14	1								

From Table ⌈1⌋ and Table ⌈2⌋, we can further generate a figure to verify the previous output figure, named Figure ⌈2⌋.

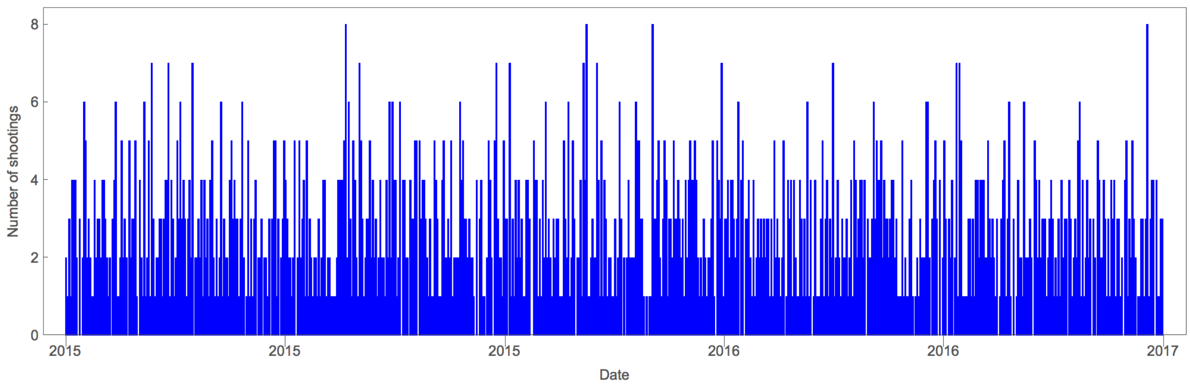


Figure 2: Daily Fatal Police Shooting Data in Table ⌈1⌋ and ⌈2⌋ by Mathematica

4 Different Reasons For Police To Shoot

There are many reasons for police to shoot at people. After statistical analysis, the numbers of each shooting cases and the corresponding proportions in US in 2015-2016 can be shown in Table ⌈3⌋.

Table 3: Reasons of being shot by the US Police in 2015-2016

Armed	*Rank*	*Number* 2015 − 2016	%
Gun	1	1074	54.96
Knife	2	289	14.79
Unarmed	3	142	7.27
Vehicle	4	119	6.09
Other	5	116	5.94
Undetermined	6	91	4.66
Toy weapon	7	87	4.45
Machete	8	16	0.82
Unknown weapon	9	12	0.61
Sword	10	8	0.41
Total		1954	100

From this table, it is obvious that for all the people being shot by the police, the ones armed with gun are the most part, which is over a half. The second high proportion corresponds to being armed knife.

5 Prediction Of The Number Of Shooting Per Day

Assume the number of shootings that happen in each day is random, then these number should follow a Poisson distribution. First, it is easy to calculate the mean $\hat{k} = \frac{1954}{731} = 2.673$. Therefore, the expected frequency can be calculated as

$$E_i = np_i = 731 \times \frac{e^{-\hat{k}}\hat{k}^i}{i!}, \tag{5.1}$$

where i = 0,1,2,...8. The expected and actual values of the number of days with 0,1,2,...,8 shootings in US in 2015-2016 can be shown in Table ⌈4⌋ and Figure ⌈3⌋.

Table 4: Expected and observed number of days with different shootings in US in 2015-2016

Shooting No.	0	1	2	3	4	5	6	7	8
Expected	50.47	134.91	180.31	160.66	107.36	57.40	25.57	9.76	3.26
Observed	50	149	163	155	115	60	23	12	4

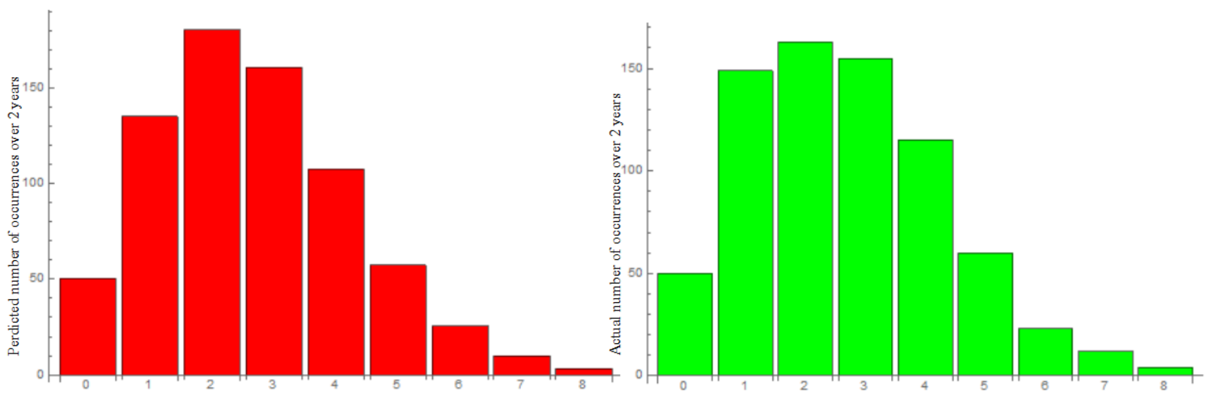


Figure 3: Expected(left) and observed(right) number of days with different shootings in US in 2015-2016

From Table ⌈4⌋, it is obvious that

$$E[X_i] = np_i \geq 1 \quad \text{for all i} = 1, ..., \text{k}, \tag{5.2}$$

$$E[X_i] = np_i \geq 5 \quad \text{for 80\% of all i} = 1, ..., \text{k}. \tag{5.3}$$

Hence, the test is

H_0 : the number of shootings per day follows a Poisson distribution with parameter $k = 2.673$.

Thus, the statistic is given by

$$X^2 = \sum_{i=0}^{8} \frac{(O_i - E_i)^2}{E_i} = 4.936. \tag{5.4}$$

Since it should follow a chi-squared distribution with $(N-1-m=7)$ degree of freedom and $X^2 < \chi^2_{0.05,7} = 14.1$, H_0 should not be rejected. Therefore, the number of shootings per day follows a Poisson distribution with parameter $k = 2.673$. Meanwhile, since $\chi^2_{0.6678,7} = 4.936$, the P-value is 0.6678.

6 Frequency Of The Shootings On Different Week Days And Months

Using Mathematica, we can create a figure analogous to Figure 3 in [2] shown as Figure ⌈4⌋ in this report.

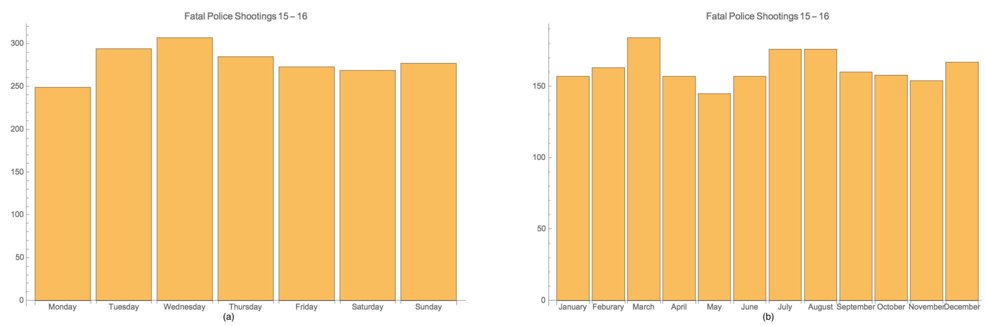


Figure 4: Frequency of the occurrence of fatal police shootings in the US on (a) different days and (b) different months from 2015 to 2016

To find out if the average number of police shootings depends on the weekday, we want to test whether the data of fatal police shootings in the US from 2015 to 2016 conform to a discrete distribution on $\Omega = \{1, 2, ..., 7\}$ that the average number of police shootings does not depend on the week day at a level of significance $\alpha = 0.05$. As January 1^{st}, 2015 is Thursday, and Year 2016 has 366 days. Thus, we can get to know that there are 104 Sundays, Mondays, Tuesdays, Wednesdays and 105 Thursdays, Fridays, Saturdays in the 731 days. We test

$$H_0: \quad \text{The data follows a multinomial distribution with parameters}$$
$$(p_1, p_2, p_3, p_4, p_5, p_6, p_7) = (\frac{104}{731}, \frac{104}{731}, \frac{104}{731}, \frac{105}{731}, \frac{105}{731}, \frac{105}{731}, \frac{104}{731})$$

where p_i(i=1,2,...,7) represents the probability that the fatal police shootings occur at the i^{th} day of the week(start from Monday). The data of fatal police shootings in the US from 2015 to 2016 is shown as Table ⌈5⌋$(E_i = np_i)$.

Table 5: "Observed" frequencies O_i and "Expected" Frequencies E_i of fatal police shootings in the US from 2015 to 2016 on different days

i	1	2	3	4	5	6	7	n
O_i	249	294	307	285	273	269	277	1954
E_i	$\frac{203216}{731}$	$\frac{203216}{731}$	$\frac{203216}{731}$	$\frac{205170}{731}$	$\frac{205170}{731}$	$\frac{205170}{731}$	$\frac{203216}{731}$	1954

Then

$$\sum_{i=1}^{7} \frac{(O_i - E_i)^2}{E_i} = 7.7368 \tag{6.1}$$

This statistic follows a Chi-square distribution with 7-1=6 degrees of freedom. As

$$\chi^2_{0.05,6} = 12.6 \quad \text{and} \quad 7.7368 < \chi^2_{0.05,6} \tag{6.2}$$

We can know that P-value of this test is larger than 5%. Therefore, we cannot reject H_0, which means there is not enough evidence showing that the average number of police shootings depends on the weekday.

7 Prediction Of The Gap Width Between Shootings

7.1 First Method: Exponential Distribution

Consider the data from January 1^{st} 2015 to December 31^{st} 2016, altogether 731 days and 1954 shootings [1]. Directly obtained from the data, there are altogether $N = 681$ intervals, 632 intervals with 0 consecutive days without shootings, 48 intervals with 1 consecutive days without shootings, 1 interval with 2 consecutive days without shootings, and 0 interval with 3 or more consecutive days without shootings, (0 consecutive days means shooting happens on successive two days).

For the first approach, time is considered as a continuum, so the Poisson distribution of the number of shootings in one day with parameter $k = 2.673$ implies that the time interval between shootings follows an exponential distribution with parameter β.

The average number of the number of shootings in one day

$$E[X] = k = 2.673 = \lambda t, \tag{7.1}$$

where λ represents the number of shootings in unit time (rate of arrivals) and t equals one day. As a result,

$\lambda = k = 2.673$, where $\beta = \frac{1}{\lambda}$ is the parameter of exponential distribution.

Since the time interval between shootings follows an exponential distribution with parameter β, the probability that the shooting happens on X^{th} day is demonstrated as

$$\begin{aligned} f_X(x) &= \frac{1}{\beta} e^{-\frac{x}{\beta}} && (7.2) \\ &= \lambda e^{-\lambda x}, && (7.3) \end{aligned}$$

in the case of U.S. shootings, $\lambda = 2.673$.

As we assume time to be continuous, p_n, the probability that the interval between two shootings is n days is equivalent to $P[n \leq X < n+1]$,

$$p_n = P[n \leq X < n+1] = \int_n^{n+1} \lambda e^{-\lambda x} = e^{-\lambda n} - e^{-\lambda(n+1)}. \tag{7.4}$$

Equation ⌈7.4⌋ is actually a density function since $p_n > 0$ for all $n \in \mathbb{N}$ and

$$\sum_{n=0}^{+\infty} p_n = \sum_{n=0}^{+\infty} e^{-\lambda n}(1 - e^{-\lambda}) = 1. \tag{7.5}$$

According to Equation ⌈7.4⌋, the probability that the interval between two shootings is n days can be calculated and is listed in Table ⌈6⌋, where n is the interval length. E_n is the expected number of the intervals with interval length n days between shootings, calculated as

$$E_n = N \times p_n = 681 \times p_n, \tag{7.6}$$

and O_n is the corresponding observed number of the intervals.

Table 6: Frequency, expected and observed number of intervals for different n

n	frequency	E_n	O_n
0	0.930955	633.98	632
1	0.064278	43.773	48
2	0.004438	3.0223	1
3	0.000306	0.2087	0
4	0.000021	0.0144	0

Figure ⌈5⌋ and ⌈6⌋ shows the expected and observed distribution of n successive days without shooting.

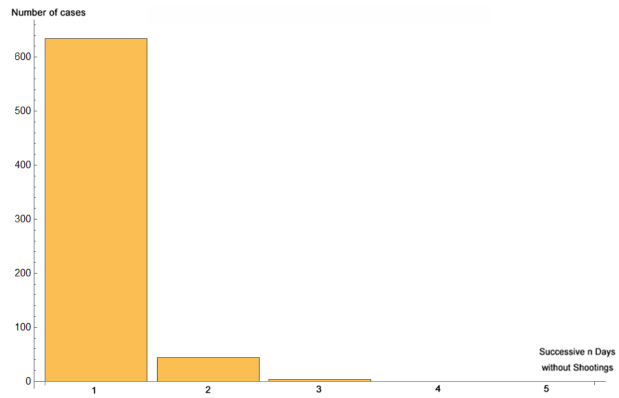


Figure 5: The expected number of intervals with consecutive n days no shootings

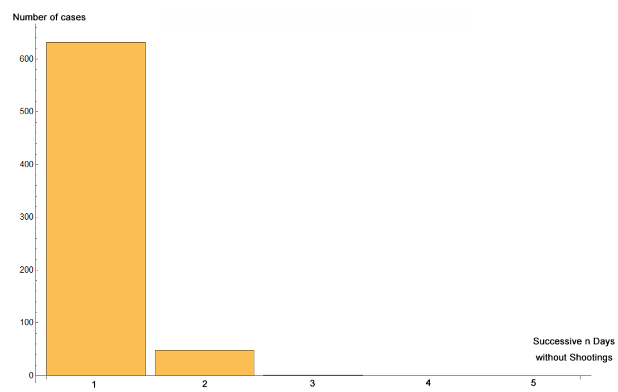


Figure 6: The observed number of intervals with consecutive n days no shootings

7.2 Second Method: Geometric Distribution

Consider the data from January 1^{st} 2015 to December 31^{st} 2016, altogether 731 days and 1954 shootings [1]. Directly obtained from the data, there are altogether N = 681 intervals, 632 intervals with 0 consecutive days without shootings, 48 intervals with 1 consecutive days without shootings, 1 interval with 2 consecutive days without shootings, and 0 interval with 3 or more consecutive days without shootings.

For the second approach demonstrated in b), discrete days are considered. There are 50 days that no shootings happen, so p_0, the probability of no shooting in one day can be derived as

$$p_0 = \frac{50}{731} = 0.0684, \tag{7.7}$$

since there are altogether 731 days measured.

Then whether shooting occurs on a given day becomes a Bernoulli trial with parameter p = 0.0684, and we can use a geometric distribution to describe the distribution of the successive days for no shooting, . The probability of no shootings on n successive days is

$$P[n] = p^n(1-p). \tag{7.8}$$

Equation ⌈7.8⌋ is actually a density function since $p_n > 0$ for all $n \in \mathbb{N}$ and

$$\sum_{n=0}^{+\infty} P[n] = \sum_{n=0}^{+\infty} p^n(1-p) = \frac{1}{p-1} \times (p-1) = 1. \tag{7.9}$$

According to Equation ⌈7.8⌋, the probability that the interval between two shootings is n days can be calculated and is listed in Table ⌈7⌋, where n is the interval length. E_n is the expected number of the intervals with interval length n days between shootings, calculated as

$$E_n = N \times P[n] = 681 \times P[n], \tag{7.10}$$

and O_n is the corresponding observed number of the intervals.

Table 7: Frequency, expected and observed number of intervals for different n

n	frequency	E_n	O_n
0	0.931600	634.42	632
1	0.063721	43.394	48
2	0.004359	2.9682	1
3	0.000298	0.2030	0
4	0.000020	0.0139	0

Figure ⌈7⌋ and ⌈8⌋ shows the expected and observed distribution of n successive days without shooting.

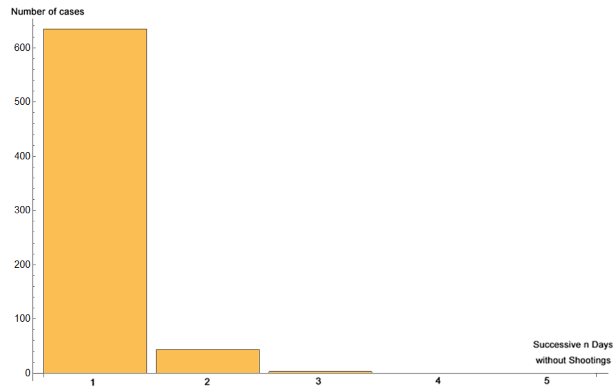


Figure 7: The expected number of intervals with consecutive n days no shootings

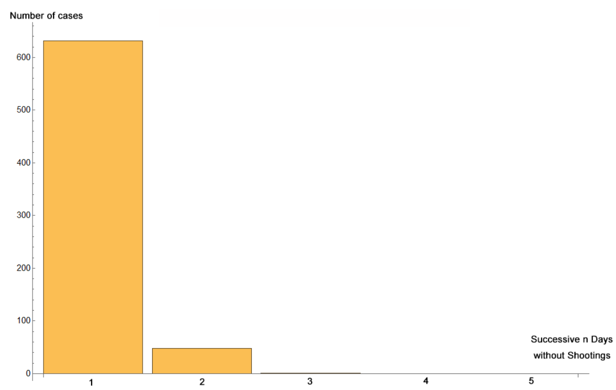


Figure 8: The observed number of intervals with consecutive n days no shootings

7.3 Comparisons Between Two Approaches

Table [8] illustrates the expected distribution of n consecutive days without shootings, obtained with two approaches respectively. E_{n_1} refers to the exponential distribution while E_{n_2} refers to the geometric distribution.

Table 8: Expected and observed distribution for two methods

n	E_{n_1}	E_{n_2}	O_n
0	633.98	634.42	632
1	43.773	43.394	48
2	3.0223	2.9682	1
3	0.2087	0.2030	0
4	0.0144	0.0139	0

Here we cannot apply the chi-squared goodness-of-fit test since if we want to satisfy the criteria that

$$E[X_i] = np_i \geq 1 \qquad \forall\ i = 1, 2, ..k, \tag{7.11}$$

$$E[X_i] = np_i \leq 5 \qquad 80\%\ of\ i = 1, 2, ..k, \tag{7.12}$$

we will get k = 2. Then the statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

will only have $k - 1 - m = 0$ degrees of freedom, where $k = 2$ and $m = 1$ since we estimate λ for exponential distribution and p for geometric distribution respectively. As a result, we cannot test the fitness of the model we construct with chi-squared goodness-of-fit test. But by observation, we can see both two methods make a good approximation.

8 The Fitting Of The Cumulative Number Of Murders

In this part, we will separate the data into two groups, in 2015 and 2016, which contain 341 and 340 days respectively. The figure of the cumulative number of murders of these two years ($m(d)$) is as Figure [9].

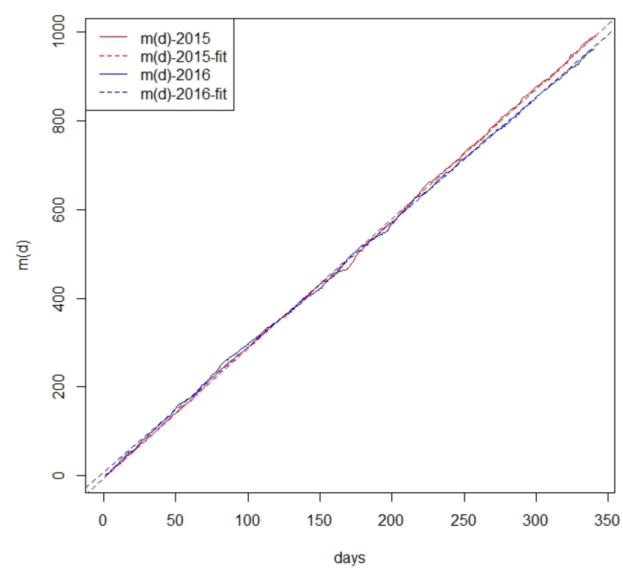


Figure 9: The cumulative number of mass shootings for each year

From the figure, $m(d)$ for these two years are both very close to straight lines. Besides, the figure indicates that the difference of the two graphs becomes more significant in the end of the year. The two lines in the figure have a tendency of separation as the x coordinate becomes larger.

After linear regression, the expressions are

$$\text{Year 2015: } m(d) = 2.921d - 4.904 \tag{8.1}$$

$$\text{Year 2016: } m(d) = 2.820d - 7.529 \tag{8.2}$$

The main statistics of the two linear regression is summarized in Table [9]

Table 9: Summary of coefficients estimates

Year	Terms	Estimate	sd.error	T.value	P.value
2015	Intercept	-4.90	0.58	-8.42	$1.06{\cdot}10^{-15}$
	Days	2.92	0.003	989.8	$< 2 \cdot 10^{-16}$
2016	Intercept	-7.52	0.57	13.28	$< 2 \cdot 10^{-16}$
	Days	2.82	0.003	978.79	$< 2 \cdot 10^{-16}$

References

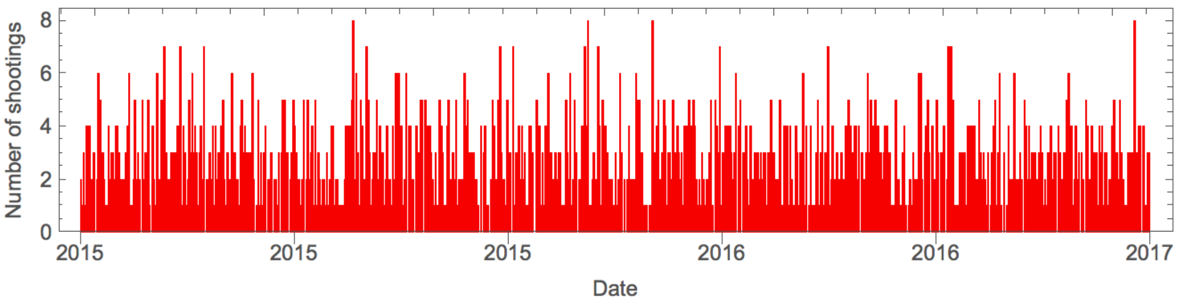
[1] The Washington Post. Fatal force. https://www.washingtonpost.com/graphics/national/police-shootings-2016/. Web. Accessed February 16th, 2017.

[2] D. Spiegelhalter and A. Barnett. London murders: a predictable pattern? *Significance*, 6(1):58, 2009. http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2009.00334.x/abstract [Online; accessed 5-July-2015].

Appendix

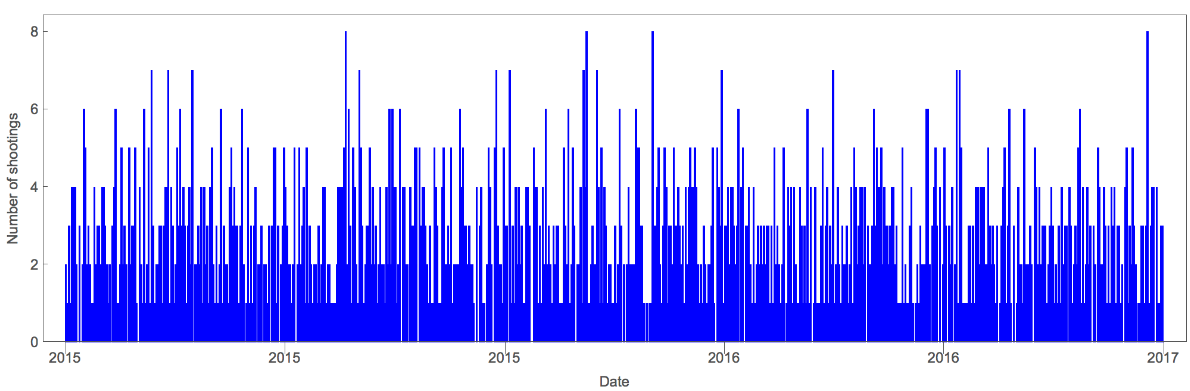
Code for Section [3.1]

```
{a} = Import["\~/Desktop/4.xlsx"];
b = Flatten[Table[a[[i, 1]], {i, 1, Length[a]}]];
DateHistogram[b, "Day", AspectRatio -> 1/5, ChartStyle -> Red,
Frame -> True, FrameLabel -> {"Date", "Number of shootings"},
DateTicksFormat -> {"Year"}, Ticks -> {Automatic, {0, 2, 4, 6, 8}},
BaseStyle -> {FontSize -> 12}]
```



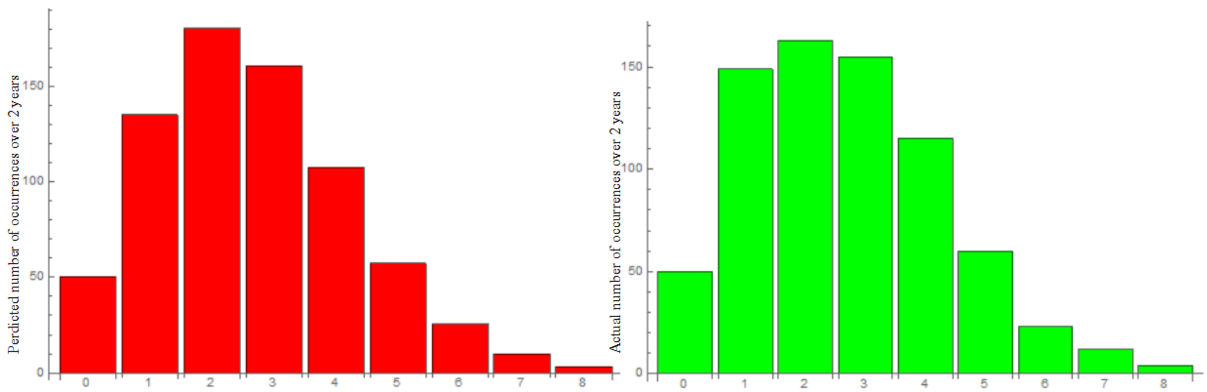
Code for Section [3.2]

```
{a} = Import["~/Desktop/3.xlsx"];
b = Flatten[Table[Table[a[[i, 1]], a[[i, 2]]], {i, 1, Length[a]}]];
DateHistogram[b, "Day", AspectRatio -> 1/3.5, ChartStyle -> Blue,
Frame -> True, FrameLabel -> {"Date", "Number of shootings"},
DateTicksFormat -> {"Year"},
FrameTicks -> {{{0, 2, 4, 6, 8}, None}, {Automatic, None}},
BaseStyle -> {FontSize -> 12}]
```



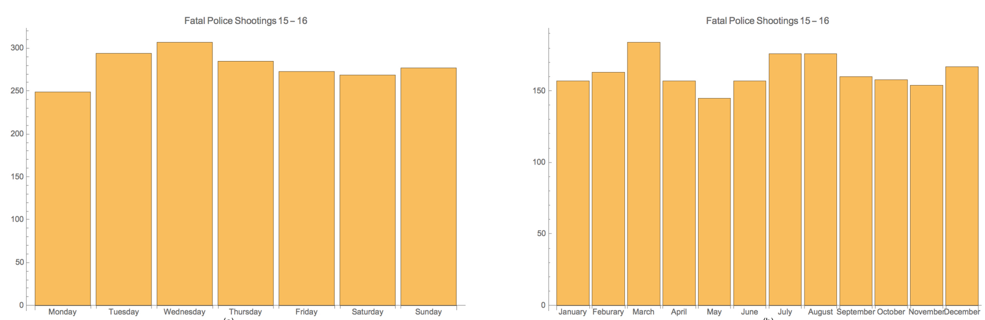
Code for Section [5]

```
BarChart[{50, 149, 163, 155, 115, 60, 23, 12, 4},
ChartLabels -> {"0", "1", "2", "3", "4", "5", "6", "7", "8"},
ChartStyle -> {Green}]
BarChart[{50.47, 134.91, 180.31, 160.66, 107.36, 57.40, 25.57, 9.76, 3.26},
ChartLabels -> {"0", "1", "2", "3", "4", "5", "6", "7", "8"},
ChartStyle -> {Red}]
```



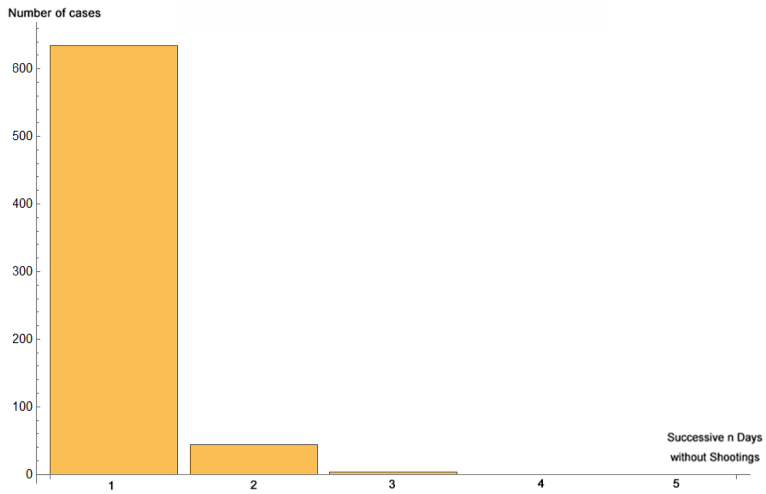
Code for Section [6]

```
h1 = BarChart[{249, 294, 307, 285, 273, 269, 277},
ChartLabels -> {"Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
"Saturday", "Sunday"},
PlotLabel -> HoldForm[Fatal Police Shootings 15 - 16]]
h2 = BarChart[{157, 163, 184, 157, 145, 157, 176, 176, 160, 158, 154, 167},
ChartLabels -> {"January", "February", "March", "April", "May", "June",
"July", "August", "September", "October", "November", "December"},
PlotLabel -> HoldForm[Fatal Police Shootings 15 - 16]]
GraphicsRow[{h1, h2}]
```

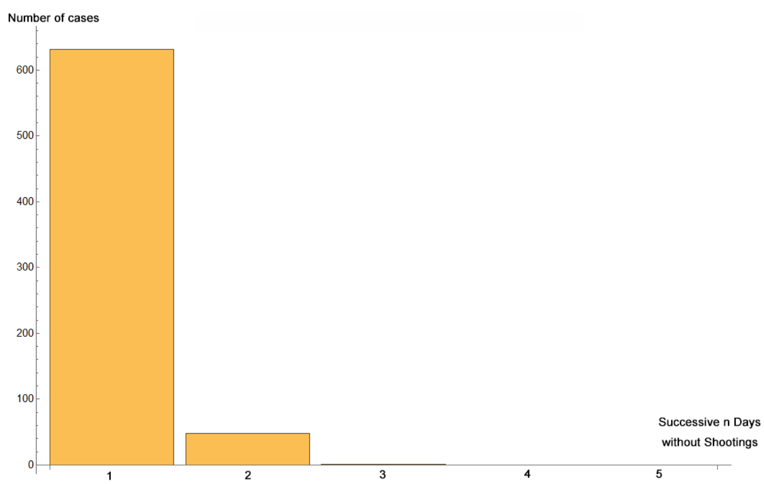


Code for Section [7.1] and [7.2]

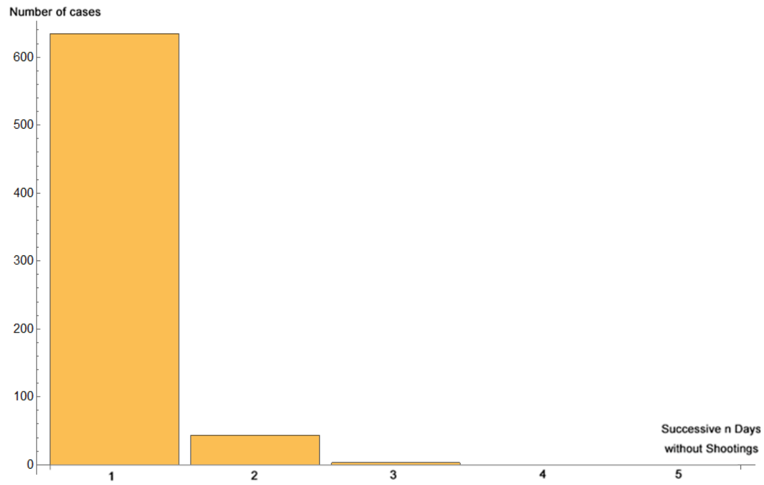
```
BarChart[{633.98, 43.773, 3.0223, 0.2087, 0.0144},
LabelStyle -> {FontFamily -> "Arial", 15, GrayLevel[0]}]
```




```
BarChart[{632, 48, 1, 0, 0},
LabelStyle -> {FontFamily -> "Arial", 15, GrayLevel[0]}]
```

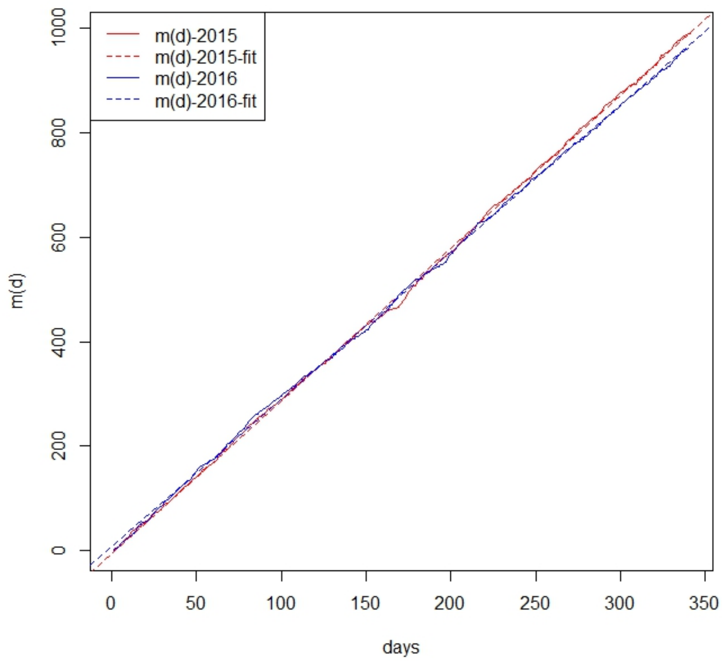



```
BarChart[{634.42, 43.392, 2.9682, 0.2030, 0.0139},
LabelStyle -> {FontFamily -> "Arial", 15, GrayLevel[0]}]
```



Code for Section [8]

```
# install.packages('dplyr')
library(dplyr)
dat <- read.csv('fatal-police-shootings-data.csv', header=TRUE)
dat <- dat$date %>% table %>% data.frame
colnames(dat) <- c('date', 'md')
dat[, 1] <- dat$date %>% as.character
n1 <- which(dat$date == '2015-12-31')
n2 <- which(dat$date == '2016-12-31')
dat2015 <- data.frame(days=1:n1, md=cumsum(dat[1:n1, 2]))
dat2016 <- data.frame(days=1:(n2-n1), md=cumsum(dat[(n1+1):n2, 2]))
lm15 <- lm(md~days, data=dat2015)
lm16 <- lm(md~days, data=dat2016)
summary(lm15)
summary(lm16)
# Plot m(d)~days
plot(dat2015$md, type='l', col='red', xlab='days', ylab='m(d)', lty=1)
abline(lm15, col='red', lty=2)
lines(dat2016$md, col='blue', lty=1)
abline(lm16, col='blue', lty=2)
legend('topleft', c('m(d)-2015', 'm(d)-2015-fit', 'm(d)-2016', 'm(d)-2016-fit'),
       col=c(rep('red', 2), rep('blue', 2)), lty=c(1, 2, 1, 2))
```


m(d)-2015
m(d)-2015-fit
m(d)-2016
m(d)-2016-fit
m(d)
days
0
200
400
600
800
1000
0
50
100
150
200
250
300
350