

De Verenigde Staten van Vuurwapengeweld

DAV Technisch Rapport
Dataset: Gun Violence

SEBASTIAAN AFLAKI

ILJA VAN IPENBURG

ELISHA NIEUWBURG

JOCHEM VAN OORSCHOT

28 juni 2018

Inleiding

Het is algemeen bekend dat Amerika het land van de geweren is. Vaak genoeg is er in het nieuws een verhaal te horen over een schietpartij in de Verenigde Staten (VS). Maar deze berichten geven niet het totaalbeeld van het vuurwapengeweld in de VS. Door de visualisatie van de data over vuurwapengeweld in de VS zijn verbanden te zien die wellicht een beter beeld kunnen scheppen van de situatie. Om dit beeld te kunnen scheppen moet er gekeken worden naar het verschil in vuurwapengeweld tussen verschillende steden, staten en jaren.

De 52 verschillende staten van Amerika zijn bijvoorbeeld niet allemaal even veilig op het gebied van vuurwapengeweld. Zo heeft de ene staat andere geweer wetten dan de andere staat [1]. De vraag is of deze wetten invloed hebben op het dodental in de staat. Volgens hebben staten met striktere wetgeving aangaande geweren minder wapengeweld. Zo wordt er verwacht dat de staten met de meest strikte wetten over geweren het laagste doden tallen hebben. Ook is het interessant om te kijken of het aantal doden en gewonden per jaar naar aanleiding van vuurwapengeweld erg verschilt. Zit er een stijgende lijn in, of daalt het juist? En wat is het verband tussen het aantal doden en gewonden? Verwacht wordt dat er een positief lineair verband bestaat tussen het aantal doden en gewonden per jaar, als dit het geval is zal dit te zien zijn bij het plotten van de data.

Veel van de grootste schietpartijen in de afgelopen jaren hebben één ding gemeenschappelijk, er is een AR-15 bij gebruikt. De AR-15 is een semi-automatisch aanvalsgeweer die zeer dodelijke kogels afschiet. Orlando, Sutherland Springs en Las Vegas zijn allemaal namen van steden waar de AR-15 is gebruikt bij een grote schietpartij [6]. Dit lijkt veel, maar is er ook daadwerkelijk een positieve toename van het gebruik van de AR-15 bij vuurwapengeweld? Verwacht wordt dat deze er is en dat dit te zien is bij het plotten van het aantal incidenten waarbij de AR-15 is gebruikt. Verder is het nog interessant om te kijken naar de relatie tussen dader en slachtoffer.

Vaak wordt gedacht dat vuurwapengeweld het meest gebeurd bij vreemden, maar er zijn ook veel incidenten waarbij de slachtoffers de dader wel kennen. Hoe zien de patronen van deze relaties er over de jaren heen uit? Er wordt verwacht er vooral veel vreemdelingen bij zitten en minder mensen die elkaar al kennen.

Methode

Bij het schonen van de dataset is er gekeken naar onnodige data die in dit onderzoek niet van toepassing was. Zo is er voor gekozen om de kolommen `incident_url`, `source_url`, `sources`, `congressional_district`, `n_guns_involved`, `state_house_district`, `state_senate_district` en `incident_url_fields_missing` te verwijderen uit de dataset. Daarnaast had het jaar 2013 te weinig data om een correct beeld te geven van de gehele dataset, dus alle data van het jaar 2013 is verwijderd uit de dataset. De overige data is goed

onderzocht op missende data, maar ook op typfouten en fout gebruik van hoofdletters. Er was een in-frequent gebruik van dubbele punten (":") gevonden in een aantal kolommen zoals `gun_type` en `gun_stolen`. Om die reden zijn alle ":" in de dataset vervangen door een enkele dubbele punt (","). Zo is er ook voor gekozen om alle NaN (ongevulde data) in de kolommen `gun_type` en `gun_stolen` te vervangen door 0:Unknown aangezien dat hetzelfde zegt als NaN. Dit is echter niet bij alle kolommen toegepast, omdat niet alle NaN waardes vervangen konden worden door een andere waarde. Bijvoorbeeld in de kolom `Notes` betekent de waarde NaN dat er geen belangrijke notities zijn van het incident en is die waarde dus niet van toepassing.

Na het opschonen van de data is Exploratory Data Analysis (EDA) toegepast op de dataset. EDA is een manier van het analyseren van verschillende soorten data op een visuele manier. Belangrijk is om voor elke soort data de meest passende manier van visualisatie te gebruiken om zo een duidelijk beeld te krijgen van de data.

Hieronder viel de dubbele staafdiagram over de moorden/gewonden per jaar is er via de kleinste kwadraten methode (*least-squares-method*) een voorspelling gedaan over de moorden en gewonden van de rest van 2018, deze werd vervolgens opgeteld bij de bestaande data van de eerste drie maanden. De kleinste kwadraten methode wordt gebruikt om bij een verzameling van punten in een vlak een formule voor een lijn te kunnen opstellen met een zo klein mogelijke kwadratische afwijking tot alle punten. Zo zijn de punten van de moorden/gewonden per maand van 2014 tot en met maart 2018 gebruikt om een formule te generen. De formule is berekend met behulp van de lineaire algebra *library* uit `numpy`. Dit resulteerde in de variabelen "a" en "b" voor de formule $y = ax + b$ waarmee voor de missende maanden van 2018 een voorspelling kon worden gedaan door de indexen van april tot en met december in te vullen voor de "x". Vervolgens kon dit opgeteld worden bij de bestaande data voor 2018.

Een andere belangrijke toepassing van EDA is de *Choropleth map*. Hierin is er gekeken naar het aantal doden en het aantal incidenten per staat en per stad. Daar zijn twee soorten visualisaties voor gemaakt: de *Choropleth map* en de *Bubble map* beiden van de VS. Een *Choropleth map* is een kaart dat elke staat van de VS een kleur toekent op basis van de waarde die men wilt onderzoeken. Deze soort is gebruikt voor het aantal doden per staat per 100.000 man en het aantal incidenten per staat per 100.000 man. Het aantal doden per 100.000 man wordt de *Death Rate* genoemd. Het is belangrijk om de te weergeven op een kaart, aangezien de ene staat een grotere populatie heeft dan de andere staat. De is berekend met de volgende formule:

$$\frac{\text{het aantal doden}}{\text{populatie}} \cdot 10^6 = \text{Death Rate}$$

De populatie per staat is gehaald uit een andere dataset [2]. Voor de populatie per staat is het gemiddelde van de populaties van de jaren 2014 tot en met 2017 genomen. Het jaar 2018 is niet met de data meegenomen, omdat er te weinig data van dat jaar was. De *Bubble map* is gebruikt voor het visualiseren van het aantal doden per stad en het aantal incidenten per stad. Een *Bubble map* is een kaart

dat de steden een cirkel geeft met een bepaalde kleur en grootte, die zijn beiden afhankelijk van het aantal doden. De groter de cirkel, de hoger het dodental. In de dataset waren niet alle latitudes en longitudes (lengte- en breedtegraden) aanwezig, daarom is uit een dataset met alle Amerikaanse steden [3] de lengte- en breedtegraden gehaald van alle steden zonder lengte- en breedtegraden. Er waren ongeveer 500 steden uit de dataset niet te vinden in de andere dataset, dus die zijn weggelaten uit de data die gebruikt is voor de kaarten.

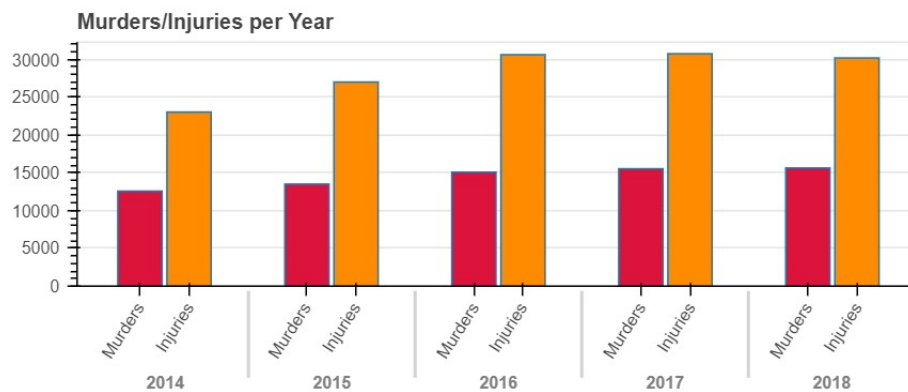
In de laatste toepassing van EDA werd er gekeken naar naar het aantal incidenten waarbij de AR-15 is gebruikt per jaar is er gekeken naar de kolommen `gun_type` en `date`. Bij de kolom `gun_type` zijn vaak meerdere typen te vinden, omdat er dan meerdere vuurwapens zijn gebruikt. Deze zijn genummerd en gescheiden door een `|`, dit is eerst gesplitst en het nummer is weggehaald, zodat er makkelijk geteld kon worden. Vervolgens werd er per jaar geteld hoe vaak de AR-15 was gebruikt bij een incident.

Hierna is dit geplot met een staafdiagram met op de x-as het jaar en op de y-as de aantallen. Er is voor een staafdiagram gekozen, omdat hierbij duidelijk het verschil in aantal tussen de jaren te zien is. Per jaar is er van een verschillende hoeveelheid incidenten bekend wat voor vuurwapen is gebruikt, hierdoor is het duidelijker om het aantal incidenten met een AR-15 te delen door het totaal aantal bekende incidenten. Door dit te plotten is het verschil per jaar nog duidelijker te zien.

Net zoals bij de data die gebruikt is bij het analyseren van het gebruik van de AR-15 zijn er bij de verschillende soorten relaties tussen dader en slachtoffer bij een incident verschillende typen te vinden. Dit is omdat er vaak meerdere slachtoffers zijn en soms ook meerdere daders. Hiervoor moesten weer per incident de verschillende relaties uit elkaar gehaald worden en moest de nummering weggehaald worden. Vervolgens is eerst het totaal van alle soort relaties geteld en daarna het soort relaties per jaar. Van het totaal van alle soort relaties is een horizontale staafdiagram geplot waarvan het logaritme van de aantallen is genomen, omdat deze ver uit elkaar liggen. Om naar de relaties per jaar te kijken is eerst per jaar geteld hoe vaak iedere relatie voorkomt. Dit is vervolgens gedeeld door het totaal aantal incidenten waarvan het type relatie bekend is. Er is gekozen om de vijf meest voorkomende relaties en *Gang vs Gang* los neer te zetten en de andere typen relaties samen te voegen onder overig. Deze data is weergegeven in een gestapelde staafdiagram, waar je duidelijk de verschillen kunt zien per type relatie.

Resultaten

Figuur 1 visualiseert de moorden en gewonden van 2014 tot en met 2018 waarbij de data van de maanden april tot en met december zijn voorspeld met de *least-squares-methode*. Uit de wapen geweld dataset blijkt dat het aantal moorden per jaar van 2014 tot en met 2018 is gestegen van 12557 moorden in 2014 naar 15612 moorden in 2018. Het aantal gewonden per jaar is van 2014 t/m 2017 gestegen van 23002 gewonden in 2014 naar 30703 gewonden in 2017. Echter zou volgens



Figuur 1: Doden en gewonden per jaar (2018 LSM)

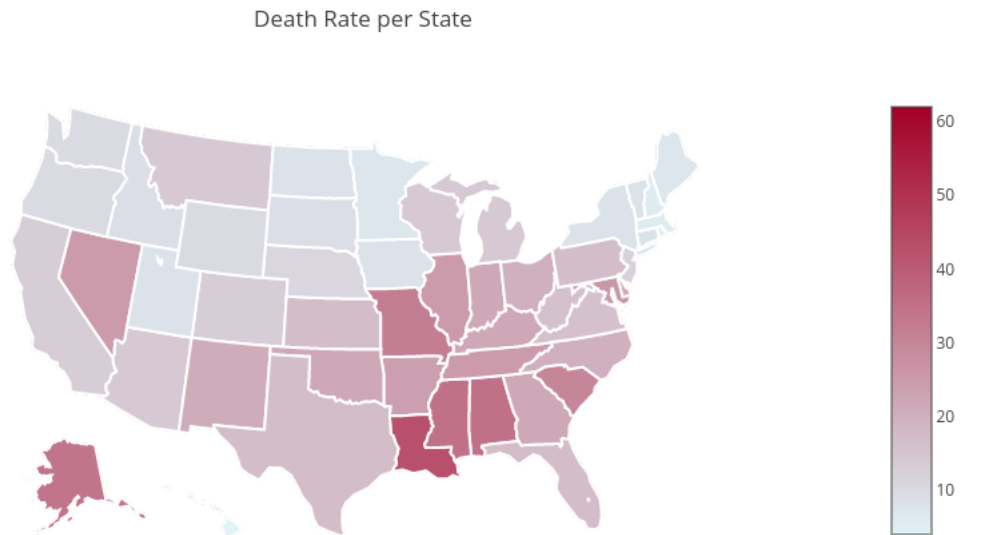
de voorspelling het aantal gewonden in 2018 ten opzichte van 2017 afnemen van 30703 naar 30147. Een ander opvallend gegeven is het verschil tussen de moorden en de gewonden per jaar, het aantal gewonden is elk jaar namelijk ongeveer het dubbele van het aantal moorden.

De *Choropleth map* toont dat de zuid-oostelijke staten van de VS een wat hogere *Death Rate* hebben dan de meer west gelegen staten. Daarnaast is Washington DC duidelijk de staat met de hoogste *Death Rate* (62), maar is de staat 'verstoppt' in de kaart (het is niet te zien op figuur 2). De volgende staat met de hoogste *Death Rate* is Louisiana met een *Death Rate* van 43. De kaart van het aantal incidenten per staat was nagenoeg hetzelfde als figuur 2, dus die is verder niet meegenomen in het onderzoek.

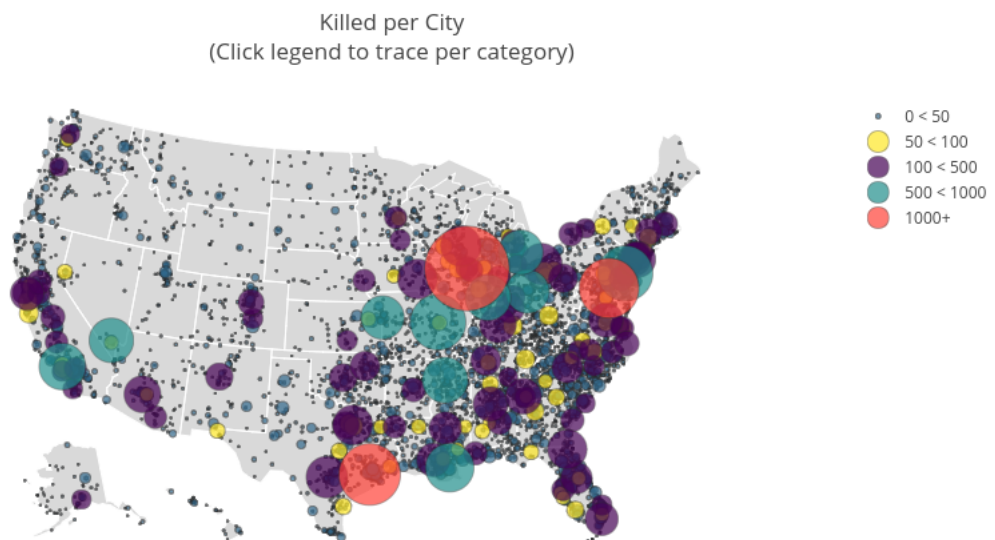
In de *Bubble map* hebben diezelfde gebieden ook de meeste cirkels. Chicago heeft het grootste dodental met 2011 doden. Houston is tweede met 1063 doden. De *Bubble map* over het aantal incidenten per stad was nagenoeg hetzelfde als figuur 3, dus die is verder niet meegenomen in het onderzoek. Daarnaast was die kaart onbetrouwbaar wegens het enorme incidenten aantal van boven de 4000 voor Chicago.

Figuur 4 komt uit het artikel *How strictly are guns regulated where you live?* van de Washington Post, 2018. Als er naast figuur 2 figuur 4 wordt gezet, is te zien dat de staten met weinig wetten zoals Alaska en Louisiana ook het meeste wapengeweld hebben. Staten met strenge wetgeving zoals Californië en Connecticut hebben ook daadwerkelijk het minste wapengeweld. Figuur 4 laat ongeveer het tegenovergestelde beeld zien figuur 2.

In figuur 5 is van 2014 tot en met 2017 een stijging te zien in aantal incidenten waarbij de AR-15 is gebruikt. 2018 heeft in deze grafiek minder incidenten waarbij de AR-15 is gebruikt, maar dit komt omdat alleen de maanden januari, februari en maart zijn meegenomen in de gebruikte dataset. Hierdoor is figuur 6 een betere representatie van het AR-15 gebruik bij vuurwapengeweld. In dit figuur is van 2014 naar 2015 een kleine daling, maar van 2015 tot en met 2018 is er een exponentiële

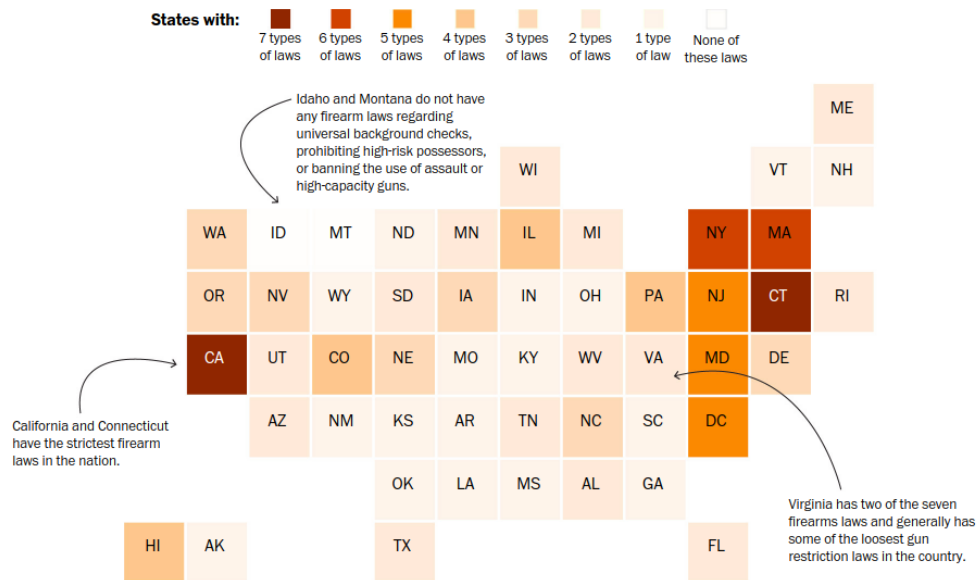


Figuur 2: *Choropleth map* over de *Death Rate* per staat

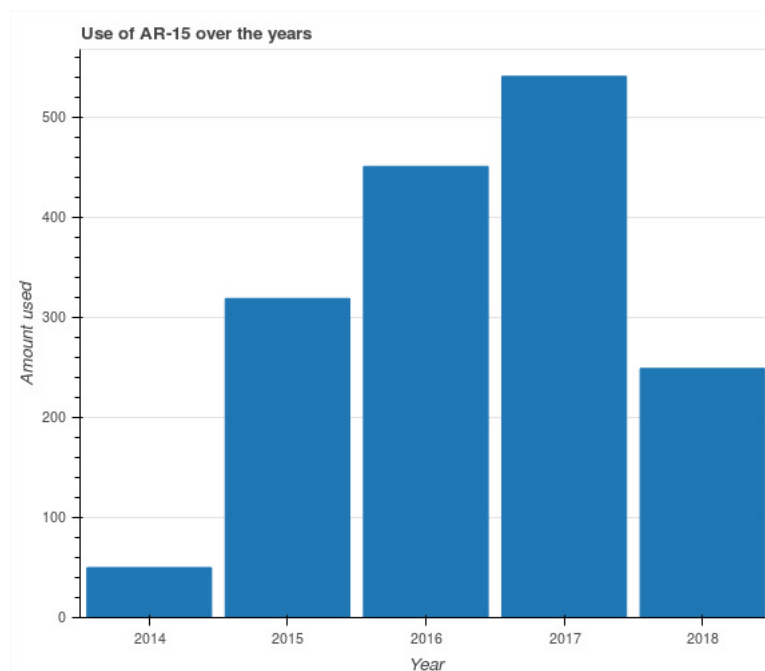


Figuur 3: *Bubble map* over het aantal doden per stad

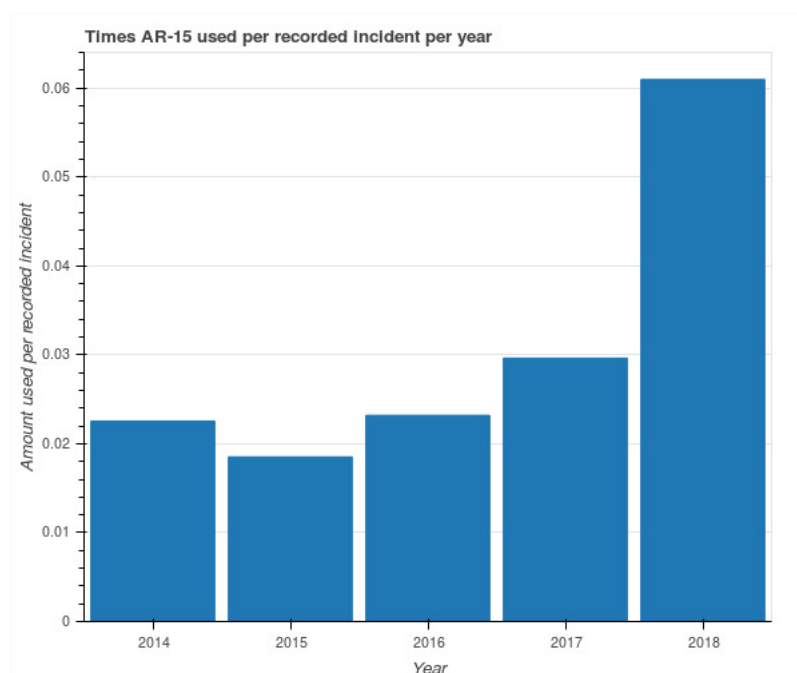
De Verenigde Staten van Vuurwapengeweld



Figuur 4: Kaart over het aantal geweer wetten per staat



Figuur 5: Aantal incidenten waarbij de AR-15 is gebruikt uitgezet per jaar. Let op: van 2018 is er alleen data van de eerste 3 maanden.



Figuur 6: Aantal incidenten waarbij de AR-15 is gebruikt per totaal aantal incidenten waarbij het wapen is vermeld per jaar.

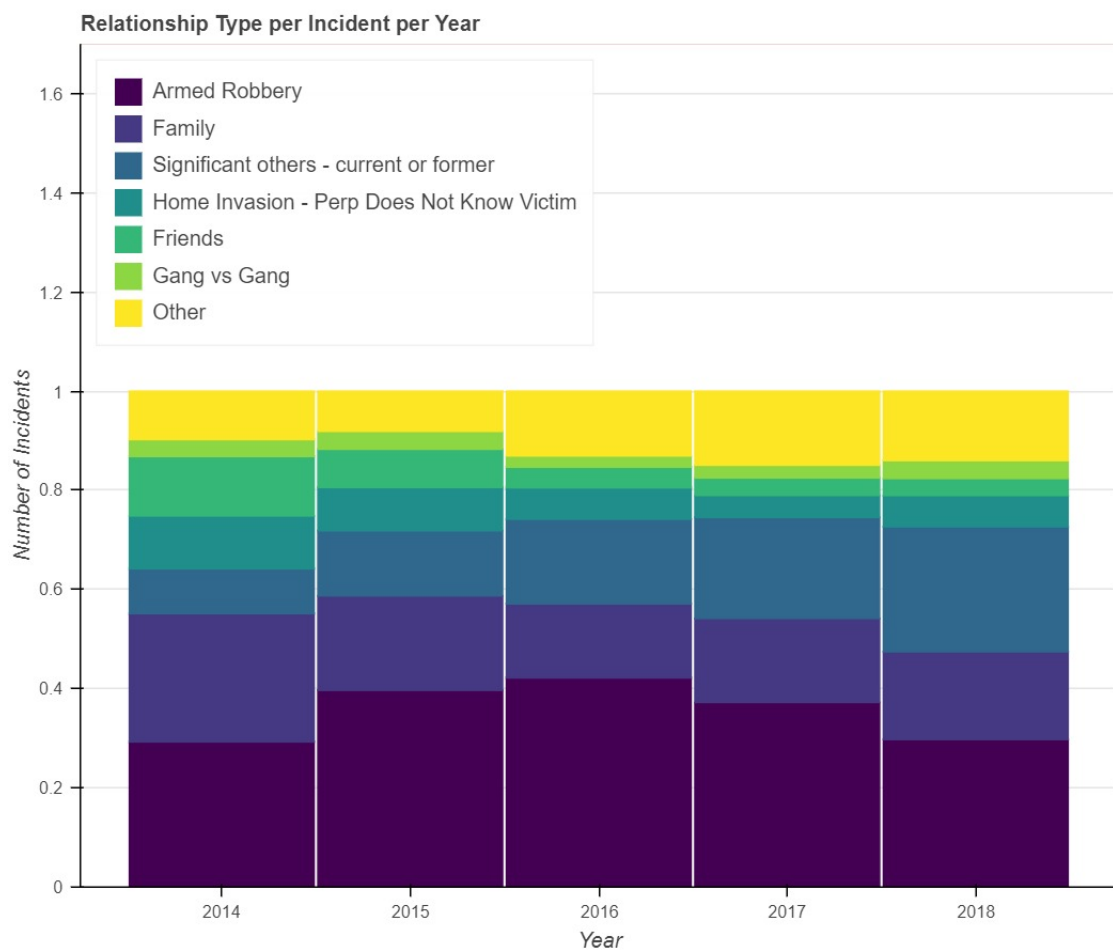
stijging, waarbij het aantal incidenten van 2017 naar 2018 meer dan verdubbelt is.

In de dataset zijn er in totaal 20.504 bekende relaties tussen de daders en slachtoffers. In figuur 7 is te zien dat hierbij de meest voorkomende relaties gewapende overval, familie en partners zijn. Gewapende overval houdt in dat de dader de overvaller is en het slachtoffer iemand die bij de overval gewond is geraakt of dood is gegaan. Dat is dan de relatie tussen de dader en het slachtoffer. Er is een significant verschil te zien tussen de nummer een en twee, maar bijna geen verschil tussen de nummer twee en drie. Ook is er te zien dat massa schiet partij het minst voorkomen van alle relaties.

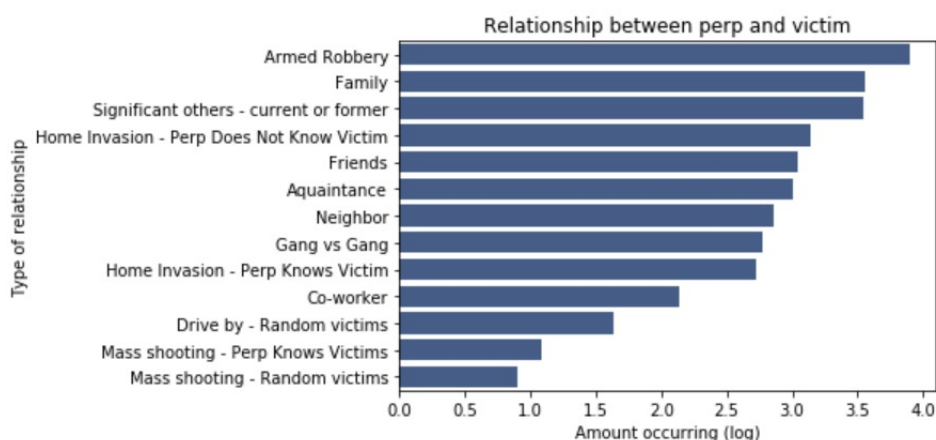
Figuur 8 laat van iedere type relatie per jaar zien hoe groot deel van het geheel aantal relaties het is. Hierdoor zie je duidelijk dat gewapende overval toeneemt tot en met 2016 en vervolgens weer daalt tot en met 2018, terwijl bij familie juist het omgekeerde gebeurt. Het aandeel van partners blijft ieder jaar toenemen en huis invasies nemen af tot 2018, waarna het weer iets toeneemt. Ook vrienden neemt af. Voor de rest is bij *Gang-vs-Gang* en overig geen duidelijke trend te zien.

Discussie

Zoals in de resultaten beschreven stijgt het dodental en gewonden tal door vuurwapengeweld elk jaar weer, wel is te merken dat deze stijging afneemt. Tussen 2014



Figuur 7: Logaritmes genomen van aantal keer dat relatie tussen dader en slachtoffer voorkomt.



Figuur 8: Hoeveelheid dat een type relatie voorkomt per jaar gedeeld door het totaal aantal incidenten waarbij de relatie is vermeld.

en 2016 stijgt het dodental met ongeveer 1000-1500 doden per jaar terwijl dit tussen 2016 en 2017 slechts rond de 500 doden per jaar ligt. Bij het gewonden tal tussen 2014 en 2016 zet zich dezelfde trend voort van een toename van ongeveer 3000 gewonden per jaar terwijl dit van 2016 op 2017 slechts rond de 200 gewonden per jaar zit. Dit wordt ook bevestigd door de voorspelling over 2018 van de kleinste kwadraten methode. Het dodental zou van 2017 op 2018 namelijk maar met ongeveer 100 stijgen en het gewonden tal daalt zelfs met zo'n 600 gewonden. Er valt natuurlijk te zeggen dat het dodental in 2018 niet zomaar te voorspellen is, maar ondanks dit vallen de resultaten van de voorspelling binnen de trend van de werkelijke data. Hieruit kan geconcludeerd worden dat het vuurwapengeweld afneemt.

Uit de resultaten van figuur 2 en 4 is te zien dat de twee kaarten tegenovergesteld zijn aan elkaar. Washington DC heeft echter een enorm hoge *Death Rate*, maar ook veel wetgeving over geweren. In figuur 3 is niet één uitschieter te zien in de staat DC, dus figuur 2 en 3 spreken elkaar op dat gebied tegen. Dit kan verklaard worden door foute berekeningen, aangezien geen rekening is gehouden met steden met dezelfde naam. Daarnaast bestaat de staat Washington en de staat Washington DC, het is niet duidelijk of dat qua dodental goed is gegaan in figuur 2. In figuur 2 is alleen de staat DC die echt de figuur 4 tegenspreekt, dus te concluderen is dat des te meer wetgeving, des te minder wapengeweld.

In vervolgonderzoek is het beter om in plaats van alleen naar steden te kijken, ook naar de staten erbij te kijken. Op die manier is er minder ruimte voor fouten door dezelfde naamgeving. In dit onderzoek was er in de gegeven tijd span geen ruimte meer om dat aan te passen.

In de inleiding werd de vraag gesteld of er een positieve toename is geweest in het gebruik van de AR-15 bij incidenten met vuurwapens. Uit de resultaten blijkt dat dit zeker het geval is, er is zelfs een exponentiële toename te zien. Wel is voor 2014 de hoeveelheid incidenten waarvan het type wapen bekend is klein vergeleken

met de andere jaren, waardoor voor dit jaar de data misschien niet volledig accuraat is. Ook was er alleen data van de eerste drie maanden van 2018, maar dit versterkt eigenlijk vooral het beeld dat er een toename is, omdat er in de drie maanden van 2018 al die helft van het totale aantal van 2017 is bereikt.

Ook werd in de inleiding de vraag gesteld hoe het patroon tussen dader en slachtoffer er over de jaren heen uit ziet. Dit is weergegeven in figuur 8. Vooral opvallend uit deze resultaten is dat het aandeel vuurwapengeweld bij partners alleen maar is toegenomen sinds 2014. Verder is het aandeel vuurwapengeweld bij familie toegenomen sinds 2016, maar heeft dit niet meer het niveau van 2014 bereikt. Voor de rest is er niet veel uit de resultaten te halen, omdat de verschillen zeer klein zijn.

Net als bij de resultaten over de AR-15 is ook bij deze resultaten weer weinig data van 2014, waardoor deze resultaten een fout beeld kunnen geven van de typen relaties van dat jaar. Ook zijn weer alleen de eerste drie maanden van 2018 bekend, waardoor deze resultaten voor 2018 misschien niet representatief zijn voor het hele jaar.

Referenties

- [1] SafeHome
<https://www.safehome.org/resources/gun-laws-and-deaths/>
- [2] *Annual Estimates of the Resident Population for the United States, Regions, States and Puerto Rico: April 1, 2010 to July 1, 2017*, (NST-EST2017-01)
<https://www2.census.gov/programs-surveys/popest/tables/2010-2017/state/totals/nst-est2017-01.xlsx>
- [3] *zip_codes_states.csv*
<https://www.gaslampmedia.com/download-zip-code-latitude-longitude-city-state-county-csv/>
- [4] Washington Post, *How strictly are guns regulated where you live?*
https://www.washingtonpost.com/graphics/2017/national/assault-weapons-laws/?utm_term=.e179d99177aa
- [5] Abigail Hauslohner
https://www.washingtonpost.com/news/national/wp/2018/04/08/feature/the-allure-of-the-ar-15-i-know-i-dont-need-it-says-one-recent-buyer-but-thats-not-the-point/?noredirect=on&utm_term=.2e266f1d218d
- [6] Matt Pearce Los Angeles Times (TNS).
<https://www.seattletimes.com/nation-world/latest-mass-shootings-all-have-ar-15-in-common/>