

# Wapengeweld in de Verenigde Staten

BSc KI 1, Data Analysis and Visualization

DAV5: Marijn Alta, Fengyuan Sun, Sarah Tol, Hugh Mee Wong

June 2018

## 1 Inleiding

*March for Our Lives*: op 24 maart 2018 gingen meer dan een miljoen demonstranten in de Verenigde Staten de straat op om te pleiten voor een strenger wapenbeleid, naar aanleiding van een aanslag op een school in Parkland, Florida. Deze schietpartij was niet de eerste in de Verenigde Staten, maar één uit de reeks van vele. Naar aanleiding van onderzoek uitgevoerd door *The Washington Post* in samenwerking met de *Federal Bureau of Investigation* (FBI) blijkt dat er sinds 1966 in de Verenigde Staten 152 van zulke schietpartijen hebben plaatsgevonden (Berkowitz, Lu, & Alcantara, 2018). Onder een schietpartij wordt hier verstaan dat minstens vier personen met een vuurwapen zijn gedood, waarbij in de telling de schutter niet wordt meegenomen. Sinds 1966 zijn er in de Verenigde Staten in totaal 1091 mensen in schietincidenten van deze aard gedood (*The Washington Post*, 2018). Schietpartijen beslaan echter niet alle vuurwapengerelateerde incidenten in de Verenigde Staten. In 2013 alleen al zijn er in totaal 11.419 doden gevallen door vuurwapengeweld (*Periscope*, 2013). Hierbij rijst de vraag wat al deze dodentallen en incidenten met elkaar gemeen hebben en waar de verschillen te vinden zijn: welke patronen zijn te vinden bij incidenten waarbij schietwapens betrokken zijn?

Om deze vraag te kunnen beantwoorden, is het van belang te kijken naar de relatie tussen staten dan wel steden onderling door de jaren heen. Hoe ziet het patroon tussen slachtoffer/dader eruit in deze incidenten? Is er een verband tussen politieke voorkeur en het aantal schietincidenten van een staat? Zijn de staten met veel schietincidenten ook de staten met de meeste doden per incidenten? Is er een verband tussen regulering van wapens en de frequentie van schietincidenten? Een *gewelddadig schietincident* wordt gedefinieerd als een incident waar minstens één person gewond is geraakt of gedood is. Met *schietincident* wordt verwezen naar een incident waarbij een schietwapen betrokken is.

De verwachting is dat er wel degelijk sprake is van verschillen tussen staten en steden onderling en dat ook door de jaren heen een patroon kan worden ontdekt. Elke staat heeft zijn eigen wapenwetgeving, waarmee het verkrijgen van een wapen in sommige staten eenvoudiger is dan in andere. Enkele staten trachten

met strengere wapencontroles het aantal wapens in omloop te verminderen, aansluitend op de gedachte dat meer wapens leidt tot meer schietincidenten. Het verschil in criminaliteitscijfers tussen staten en steden onderling zou hieraan gerelateerd zijn.

Uit eerder onderzoek is gebleken dat 57% van de doden gevallen in schietincidenten bekend was met de dader, in een kwart van de gevallen gaat het om leden van hetzelfde gezin. Op basis van deze cijfers is de verwachting dat een relatief hoog percentage van de onderzochte data gevallen van huishoudelijk geweld betreft. Dit in tegenstelling tot de algemene opvatting dat dergelijke incidenten voornamelijk door bendeleden worden veroorzaakt.

## 2 Methode

Voor het analyseren van de data is in Python gebruik gemaakt van Jupyter Notebook. Het manipuleren van de datasets is gedaan middels de modules Numpy en Pandas, waarna met Bokeh bevindingen zijn geplot voor de visualisatie.

### 2.1 Datasets

De data over wapengerelateerde incidenten in de Verenigde Staten die hebben plaatsgevonden in de periode van 2014 tot en met maart 2018 is verkregen uit de *Gun Violence Archive* en door de auteur middels *scraping* omgezet in een csv-bestand. Over het doel van het publiceren van deze dataset schrijft de GitHub-gebruiker:

There's currently a lack of large and easily-accessible amounts of detailed data on gun violence. This project aims to change that; we make a record of more than 260k gun violence incidents, with detailed information about each incident, available in CSV format. We hope that this will make it easier for data scientists and statisticians to study gun violence and predict future trends.

De originele set bevat 29 kolommen en 239.677 verschillende incidenten, waarvan voor het beantwoorden van de hierboven genoemde vragen enkele kolommen zijn verwijderd (**tabel 1**). De set bevat incidenten van 1 januari 2013 tot en met 31 maart 2018. Hiervan zijn twee incidenten, de schietpartij in Las Vegas van 1 oktober 2017 en incident met ID 1081885, handmatig door de auteur verwijderd wegens complicaties bij het ophalen van de data. De data uit 2013, het jaar waarin de *Gun Violence Archive* werd opgezet, is met 278 geregisteerde incidenten (tegenover een gemiddelde van 56.399 van de vier daaropvolgende jaren) incompleet. De onvolledige registratie uit dit jaar bemoeilijkt het uitvoeren van een analyse op jaarbasis en is om die reden niet meegenomen in de definitieve dataset. Na verwijdering van incidenten uit het betreffende jaar beslaat de uiteindelijke dataset 239.399 incidenten van 1 januari 2014 tot en met 31 maart 2018.

Data over de populatiegrootte van de Amerikaanse staten is gebaseerd op de geschatte waarden die het *United States Census Bureau* op 1 juli 2017 vrij heeft gegeven. In datzelfde jaar publiceerde het *Bureau of Alcohol, Tobacco, Firearms and Explosives* (ATFE) een rapport over de wapenomloop in de Verenigde Staten, waarvan cijfers over het aantal geregistreerde vuurwapens per staat is opgenomen in de dataset over wapenbezit.

Voor het onderzoeken naar een mogelijk verband tussen wapengeweld en de politieke voorkeur van een staat is een dataset opgesteld met uitslagen van de presidentsverkiezingen van 2016. Per staat zijn hiervoor drie variabelen ingevuld: *democrat* (Democratisch), *republican* (Republikeins), *other* (Overig). Onder deze laatste categorie vallen alle stemmen die niet voor de Democratische of Republikeinse partij zijn uitgebracht.

## 2.2 Exploratieve Data-analyse

Exploratieve data-analyse is verdeeld in vier verschillende delen (*univariate non-graphical*, *multivariate non-graphical*, *univariate graphical*, *multivariate graphical*) die ieder een andere aanpak vereisen. Hierbij worden de genoemde vragen beantwoord aan de hand van deelvragen.

De jaarlijkse cijfers over incidenten van 2014 tot en met 2017 zijn met elkaar vergeleken om een beeld te schetsen over het verloop van het aantal incidenten op jaarbasis. Om een betere voorspelling te kunnen doen over het verdere verloop van 2018 en de daaropvolgende jaren, is de data ook op maandbasis vergeleken. Deze maandelijkse cijfers zijn uiteindelijk vergeleken met de cijfers die over de eerste drie maanden van 2018 beschikbaar zijn gesteld.

Voor het analyseren van het verschil in aantal incidenten tussen staten en steden, is na de vergelijking van de totale aantallen ook onderscheid gemaakt op basis van het aantal slachtoffers: per staat is gekeken naar het aantal schietpartijen en het aantal gewelddadige incidenten, waarbij het laatste verwijst naar incidenten waarbij minstens een persoon verwond of gedood is. Aangezien het aantal inwoners per staat aanzienlijk verschilt, zijn voor de drie typen incidenten tijdens de analyse ook de aantallen per capita meegenomen door per staat het totale aantal te delen door de populatiegrootte.

Relaties tussen slachtoffer en dader is onderzocht aan de hand van informatie in de kolom *incident\_characteristics* uit de *Gun Violence* dataset. Deze kolom beschrijft de aard van het incident aan de hand van kernwoorden met korte beschrijvingen. Deze beschrijvingen zijn echter niet specifiek genoeg om een uitspraak te doen over de directe relatie tussen dader en slachtoffer: zo is bij huishoudelijk geweld (*domestic violence*) welke leden van de familie betrokken waren bij het incident. Om deze reden is het onderzoek naar de relaties tussen betrokkenen van een incident op een hoger analyiseniveau uitgevoerd door de aandacht te richten op de *aard* van het incident, waarbij onderscheid is gemaakt tussen schietpartijen (*mass shootings*), alcoholgerelateerde incidenten (*alcohol involvement*), bendegerelateerde incidenten (*gang involvement*), huishoudelijk geweld (*domestic violence*), gewapende overvallen (*armed robbery*) en drugsgerelateerde incidenten (*drug involvement*).

Het is gebruikelijk de politieke voorkeur van een staat te baseren op de partij die het grootste aandeel van de stemmen voor zich weet te winnen. Door hierbij de overige stemmen niet in acht te nemen, wordt een significant deel van de stemmen verwaarloosd: met 40% voor de grootste partij wordt 60% van de stemmen niet meer meegenomen in de verdere analyse. Door per staat de percentages voor iedere partij tegenover het aantal incidenten weer te geven, wordt rekening gehouden met een verschil in stempatronen tussen staten met dezelfde voorkeurspartij.

### 2.3 *In-depth* Analyse

Voor het berekenen van de correlatiecoëfficiënt is gebruik gemaakt van de volgende formule:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

Correlatiecoëfficiënt  $r$  tussen twee variabelen  $x$  en  $y$  op interval- of ratioschaal wordt hierbij berekend door het quotiënt te nemen van de covariantie van de twee variabelen en het product van de standaardafwijkingen. Bij het analyseren van de resultaten is voor de berekeningen gebruik gemaakt van Pythons module *scikit-learn*. Met de berekende waarde is te bepalen of een (lineaire) relatie tussen de twee variabelen aanwezig is, waarbij men tracht de *mean squares error* te verkleinen om de betrouwbaarheid van voorspellingen te vergroten.

## 3 Resultaten

De resultaten zijn per deelvraag opgedeeld om het overzicht voor de lezer te behouden. De eerste twee vragen zijn algemeen en de laatste drie zijn specifiek gericht op de aanvullende datasets.

### 3.1 Door de jaren heen

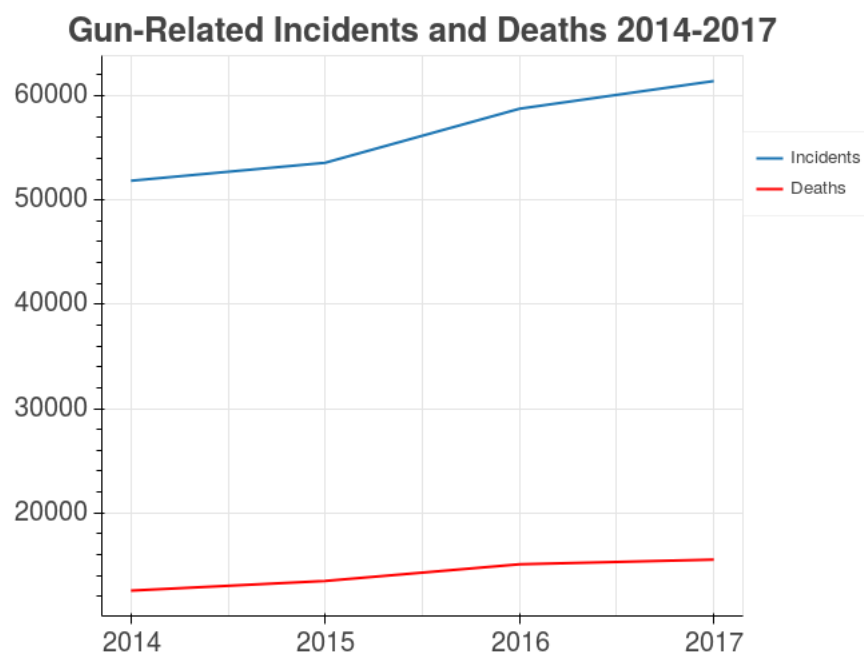


Figure 1: Het aantal vuurwapengerelateerde incidenten en doden in de Verenigde Staten

In de periode van 2014 tot en met 2017 hebben er jaarlijks gemiddeld 56399 vuurwapengerelateerde incidenten plaatsgevonden. Hierbij is het aantal incidenten ieder jaar ten opzichte van de voorgaande jaren gestegen: van 51854 in 2014 tot 61401 in 2017. Gemiddeld steeg het aantal incidenten met 106% per jaar.

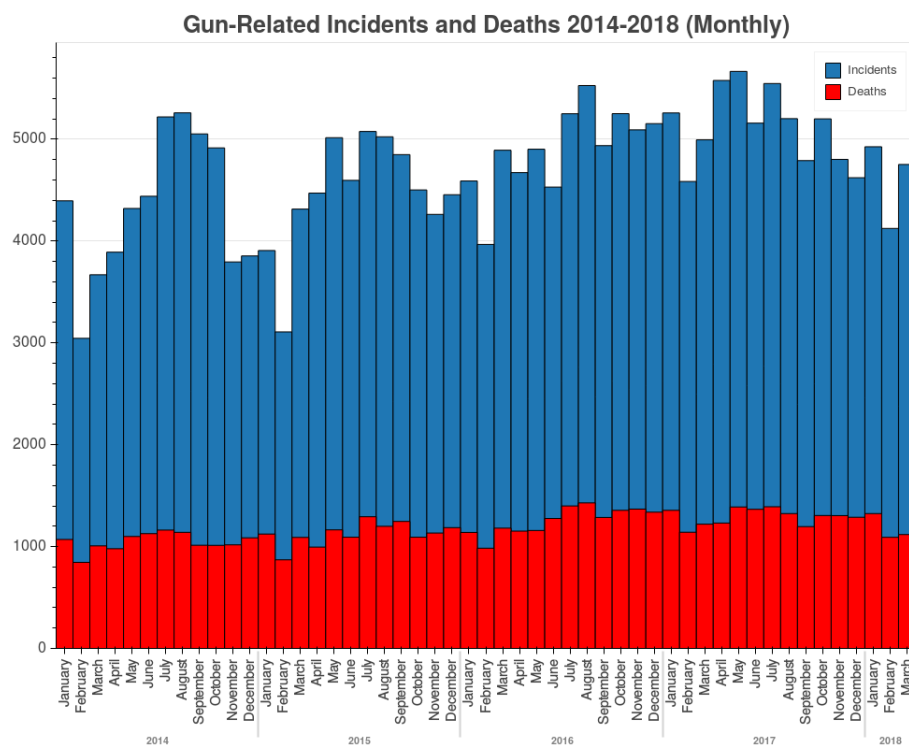


Figure 2: Per maand het aantal vuurwapengerelateerde incidenten en doden in de Verenigde Staten

In de periode van 2014 tot en met 2017 hebben er maandelijks 4694 incidenten plaatsgevonden. Februari 2014 was met 2045 incidenten in deze set de maand met de minste incidenten en het recordaantal van 5666 incidenten was in mei 2017. In 2014 tot en met 2017 hebben er in januari, februari en maart respectievelijk 4538, 3676 en 4467 incidenten gemiddeld plaatsgevonden. 2018 telde voor ieder van deze drie maanden meer incidenten: 4926 in januari, 4124 in februari en 4752 in maart. Hoewel het gemiddelde aantal incidenten en doden per jaar is gestegen, blijkt voor ieder van deze variabelen dat een piek waar te nemen is in de periode rond juni. Ook het aantal schietpartijen is per jaar niet constant: voor incidenten van deze aard is een vergelijkbare piek in de zomerperiode waar te nemen (**Figure 3**).

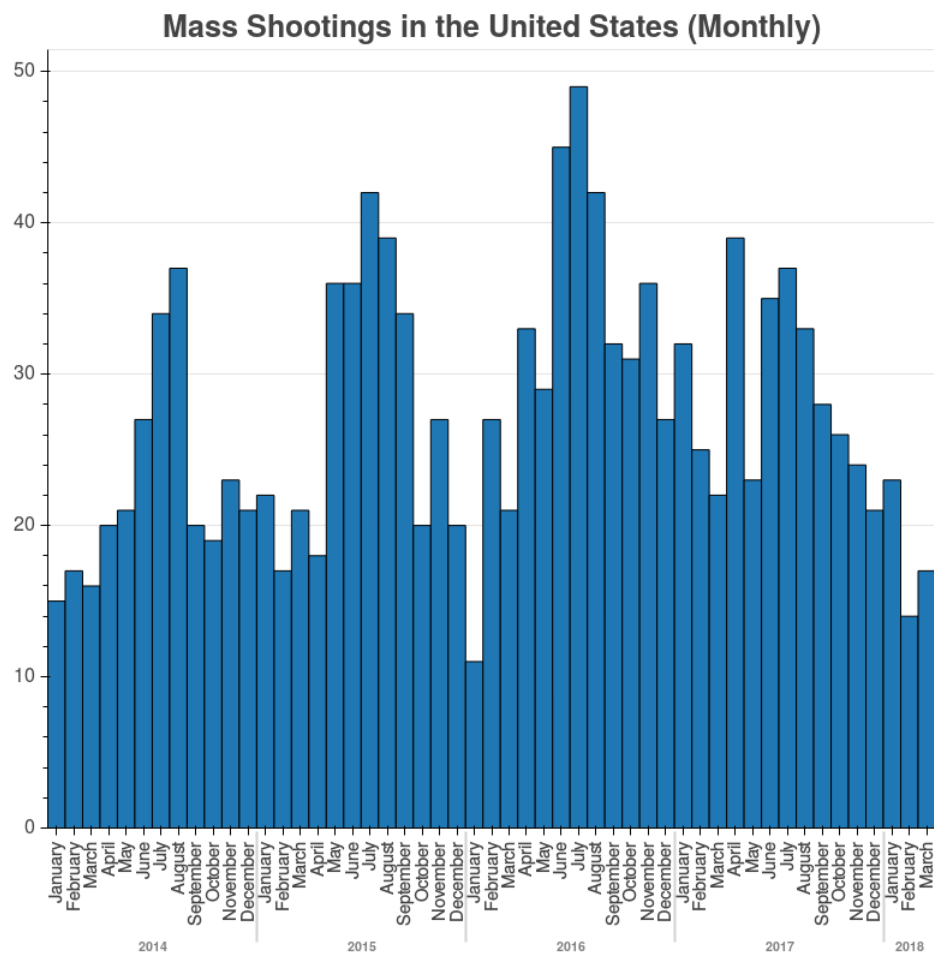
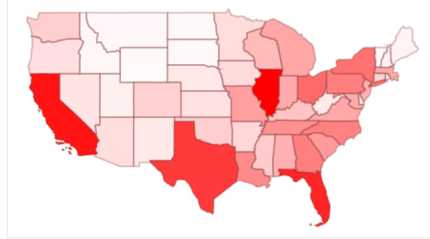


Figure 3: Per maand het aantal schietpartijen in de Verenigde Staten

### 3.2 Verschillen tussen staten

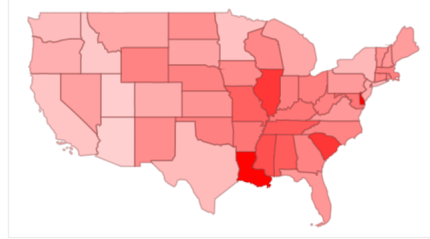
Met een gemiddelde van 4694 incidenten per staat heeft Hawaii met een gemiddelde van 289 het laagste gemiddelde en Illinois met gemiddeld 17538 incidenten het hoogste aantal incidenten.

**Incidents per state**



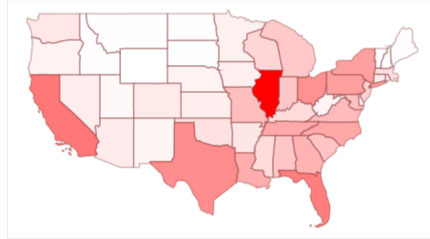
(a) Alle incidenten

**Incidents per capita**



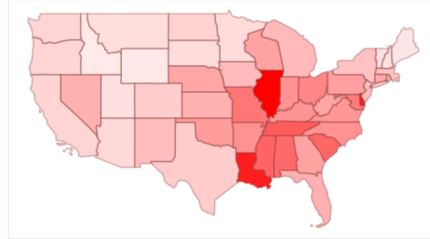
(b) Incidenten per capita

**Violent incidents per state**



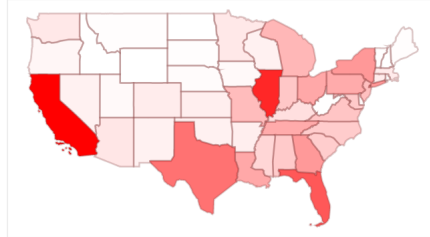
(c) Gewelddadige incidenten

**Violent incidents per capita**



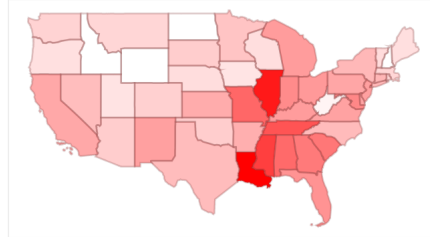
(d) Gewelddadige incidenten per capita

**Mass shootings per state**



(e) Schietpartijen

**Mass shootings per capita**



(f) Schietpartijen per capita

Figure 4: Links afgebeeld zijn kaarten gebaseerd op het totale aantal incidenten (*incidents*), geweldadige incidenten (*violent incidents*) en schietpartijen (*mass shootings*) per staat, rechts de aantallen voor elke staat per capita. Een hogere kleurintensiteit duidt op een hoger aantal incidenten in verhouding tot andere staten. Naar voorbeeld van *Gun Violence Archive* is voor de visualisatie de keuze gemaakt om Hawaii en Alaska niet weer te geven, hoewel de data van deze twee staten wel is meegenomen in de analyse.



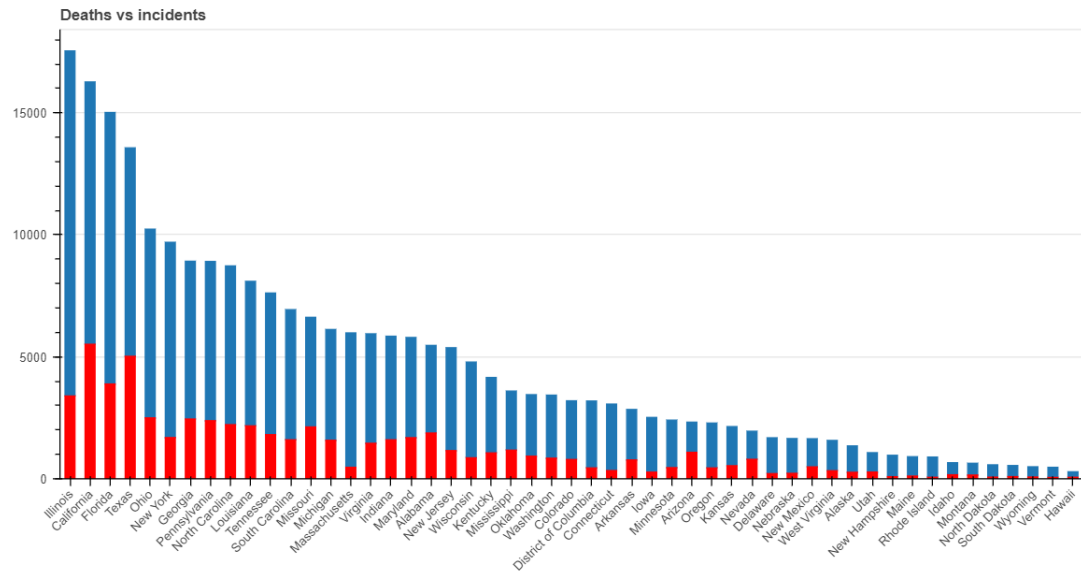
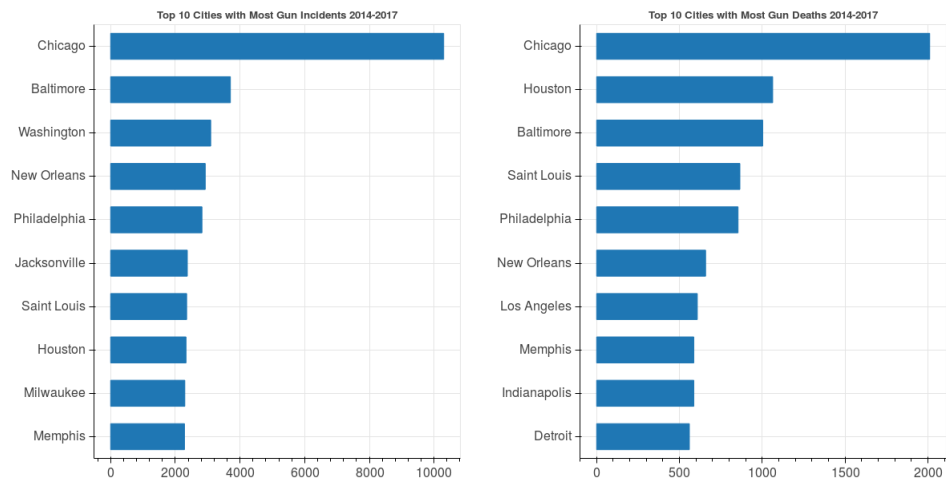


Figure 5: Per staat het aantal incidenten en het aantal doden in de periode van 2014 tot en met 2017.

### 3.3 Verschillen tussen steden

Van de vijf steden met de meeste vuurwapengerelateerde incidenten zijn er vijf steden die daarmee ook relatief veel vuurwapendoden hebben. Chicago is de stad die te kampen heeft met de meeste vuurwapengerelateerde incidenten en doden, het aantal incidenten in deze stad beslaat 67% van het totale aantal incidenten in de staat Illinois.



### 3.4 Karakteristieke incidenten

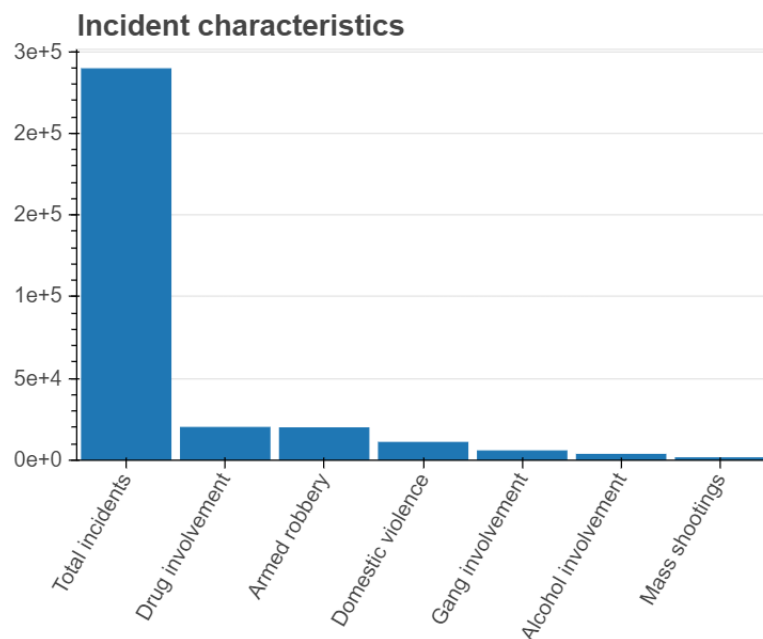


Figure 6: Het aantal incidenten gefilterd op de aangegeven kenmerken in *incident\_characteristics* uit de *Gun Violence* dataset. Ter vergelijking is links het totale aantal incidenten uit de periode van 2014 tot en met 2017 weergegeven.

In **Figure 7** is op te maken dat een toename van het totale aantal incidenten in verband staat met het aantal gewelddadige incidenten en het aantal schietpartijen. Met de populatiegrootte en het aantal incidenten is een correlatiecoëfficiënt van 0.7 berekend.

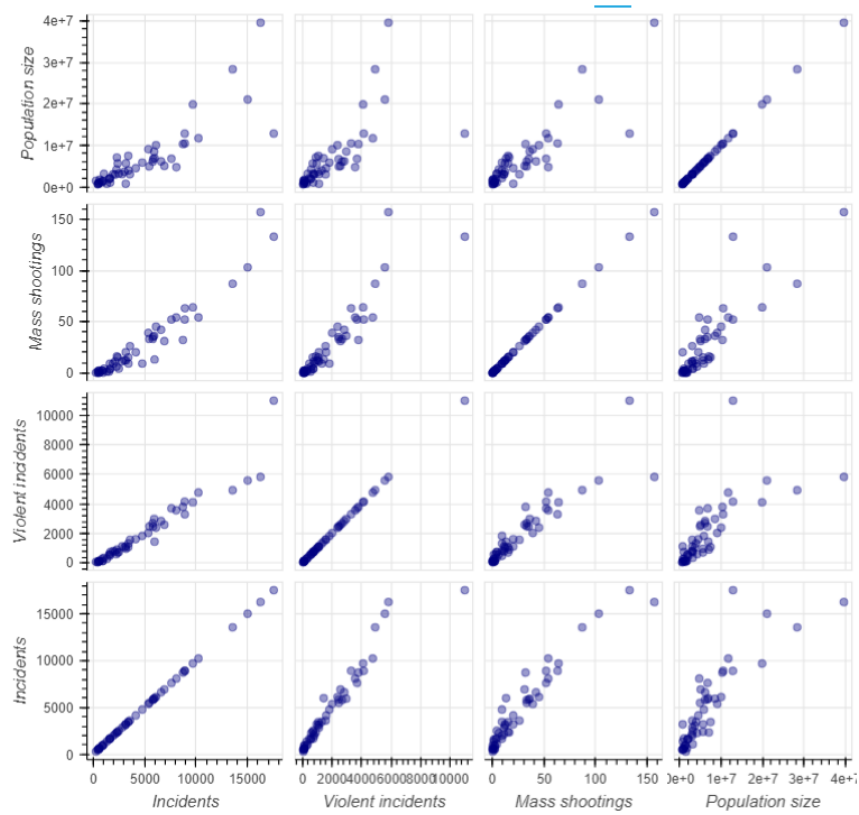


Figure 7: Het aantal incidenten, gewelddadige incidenten, schietpartijen en de populatiegrootte tegen elkaar afgezet.

### 3.5 Wapenbezit

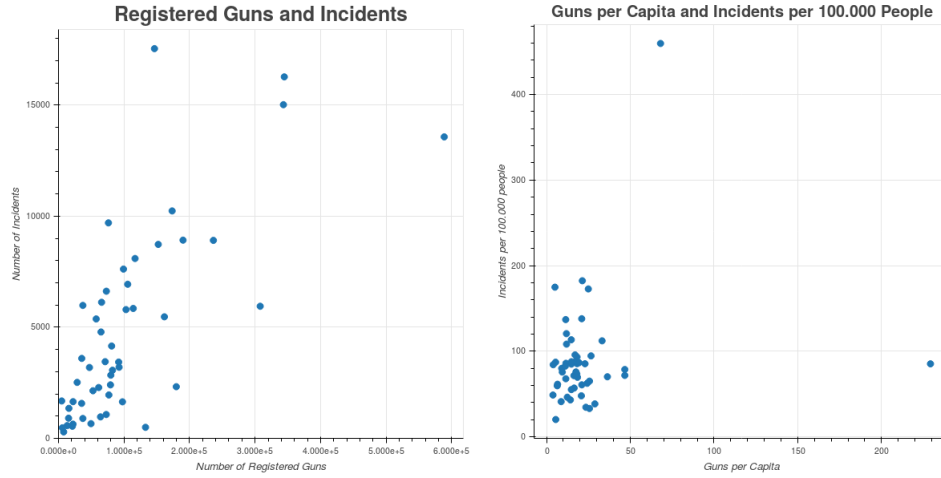


Figure 8: Links het aantal geregistreerde wapens per staat tegenover het aantal incidenten ( $r = .8$ ), rechts het aantal geregistreerde wapens tegenover het aantal incidenten per 100,000 mensen in een staat ( $r \leq .01$ ).

### 3.6 Politieke voorkeur

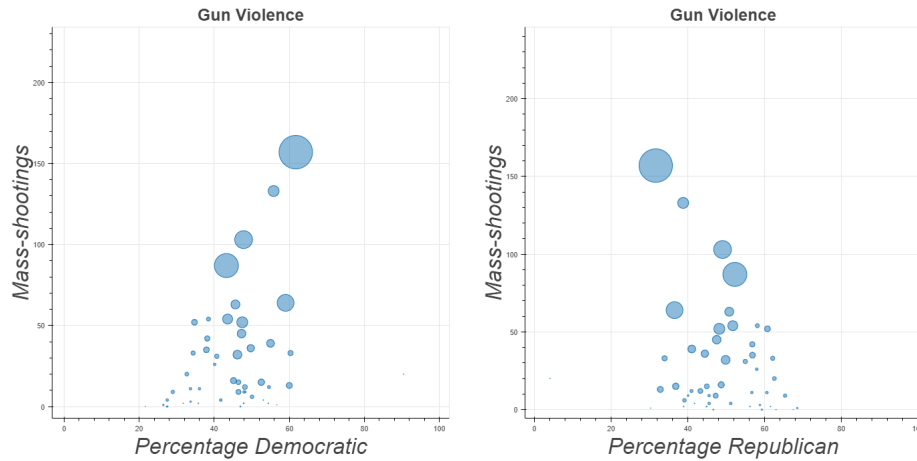


Figure 9: De neiging tot aanhang van de twee grootste politieke partijen in de Verenigde Staten tegenover het aantal schietpartijen per staat. De grootte van de cirkels geeft de grootte van de staat weer. In beide gevallen was de correlatiecoëfficiënt  $r \leq .01$

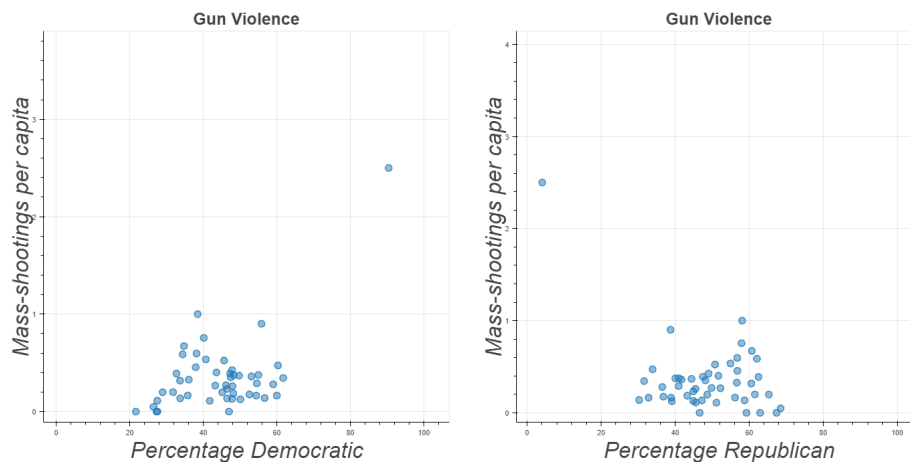


Figure 10: Links afgebeeld de cijfers over de Democratische aanhang tegenover het aantal schietpartijen, rechts de Republikeinse aanhang tegenover het aantal schietpartijen. In beide gevallen was de correlatiecoëfficiënt  $r \leq .01$

## 4 Discussie

Uit analyse van de verkregen data blijkt dat zowel het totale aantal vuurwapengerelateerde incidenten als het aantal schietpartijen in de periode van 2014 tot en met 2017 is gestegen, waarbij de verwachting is dat deze stijgende trend in 2018 door zal lopen: uit de data van de eerste drie maanden van dit jaar zijn de telwaarden hoger dan die van de eerste drie maanden in voorgaande jaren. Of het hierbij gaat om een toename van het aantal incidenten of een toename van het aantal *registraties*, is nog onduidelijk. De *Gun Violence Archive* verzamelt data door dagelijks automatisch en handmatig aanvragen te doen bij meer dan 2000 bronnen uit de media, politierapporten en bronnen van andere instanties. Vanzelfsprekend is dat een ongemeld incident niet wordt opgenomen in de database. Verder onderzoek zou zich kunnen richten op de volledigheid van zulke databases en het verschil in mate van registratie door de jaren heen, waarbij men mogelijk rekening moet houden met het toenemende verkeer op social media platforms.

Opvallend is dat in ieder van de geanalyseerde jaren een significante toename van het aantal incidenten waar te nemen is in de zomerperiode van mei tot en met augustus. Ook is er in deze periode een positieve correlatie waar te nemen tussen het totale aantal incidenten en het aantal schietpartijen: de verhouding tussen incidenten en schietpartijen blijft nagenoeg gelijk. Hoewel dit te wijten kan zijn aan een reeks van factoren, lijkt het niet overbodig om dieper in te gaan op de effecten van de temperatuurstijging op het aantal wapengerelateerde incidenten in de Verenigde Staten.

Het totale aantal geregistreerde wapens vertoont een positieve correlatie met het aantal incidenten: staten waarin meer wapens geregistreerd zijn, hebben

te kampen met significant meer vuurwapengerelateerde incidenten. Het aantal wapens per capita vertoont daarentegen geen verband met het aantal incidenten: het aantal wapens per inwoner is geen voorspellende factor voor het aantal incidenten per staat. Hoewel er rekening is gehouden met de populatiegrootte, is hierbij een mogelijk effect van populatiedichtheid niet in acht genomen. Het verschil tussen de twee bevindingen over wapenbezit leidt tot de vraag of inperking van wapenbezit effect heeft op het terugdringen van wapengeweld. Op basis van bevindingen uit de beschreven analyse zou men kunnen zeggen dat er geen sprake is van een verband tussen politieke voorkeur in een staat en het aantal schietincidenten. *Voorkeur* bleek hiervoor echter te zwak gedefinieerd. Een uitspraak over voorkeur baseren op de meerderheid van de stemmen laat daarmee een aanzienlijk deel van de stemmen buiten beschouwing. Het uitvoeren van analyses op basis van percentages per groep op zich heeft het definitieprobleem niet in zijn geheel geëlimineerd. Verkiezingsuitslagen vereisen enige flexibiliteit in de interpretatie: een stem voor een partij of een vertegenwoordiger blijkt vaak niet representatief te zijn voor de eigen waarden en idealen, waarbij het mogelijk is dat een stem *voor* voornamelijk een stem *tegen* de oppositiepartij is (Pew Research Center, 2016).

## References

- [1] Gun Violence Archive. (2018). *Standard Reports*. Geraadpleegd op <http://www.gunviolencearchive.org/reports>
- [2] Berkowitz, B., Lu, D., & Alcantara, C. (2018). *The terrible numbers that grow with each mass shooting*. The Washington Post. Geraadpleegd op [https://www.washingtonpost.com/graphics/2018/national/mass-shootings-in-america/?utm\\_term=.564c5bf7e48d](https://www.washingtonpost.com/graphics/2018/national/mass-shootings-in-america/?utm_term=.564c5bf7e48d)
- [3] Periscope. (2013). *U.S. Gun Deaths in 2013*. Geraadpleegd op <https://periscope.com/news/2013/02/thoughts-on-visualizing-u-s-gun-murders>
- [4] jamesqo op GitHub. (2018). *Gun Violence Data*. GitHub repository, geraadpleegd op <https://github.com/jamesqo/gun-violence-data>
- [5] Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2017. United States Census Bureau.
- [6] Firearms Commerce in the United States: Annual Statistical Update 2017. United States Department of Justice: Bureau of Alcohol, Tobacco, Firearms and Explosives.
- [7] Geiger, A. (2016). For many voters, it's not which presidential candidate they're for but which they're against. Pew Research Center. Geraadpleegd op <http://www.pewresearch.org/fact-tank/2016/09/02/for-many-voters-its-not-which-presidential-candidate-theyre-for-but-which-theyre-against/>

## Appendix

Table 1: Dataset wapengeweld: verwijderde kolommen

Kolom	Reden voor verwijdering
address	Per incident is informatie over de staat en stad beschikbaar, incidenten afbakenen op adres is overbodig.
incident_url	Het is niet van belang ieder incident individueel te bekijken.
source_url	Het is niet van belang ieder incident individueel te bekijken.
incident_url_fields_missing	Na verwijdering van incident_url is het ontbreken van incident_url niet meer relevant.
congressional_district	Per incident is informatie over de staat en stad beschikbaar, incidenten afbakenen op district-nummer is overbodig.
gun_stolen	Incomplete data: bij 41,5% van de incidenten is deze kolom leeg.
gun_type	Incomplete data: bij 41,5% van de incidenten is deze kolom leeg.
location_description	Incomplete data: bij 82,4% van de incidenten is deze kolom leeg.
n_guns_involved	Incomplete data: bij 41,5% van de incidenten is deze kolom leeg.
notes	Het is niet van belang ieder incident individueel te bekijken.
participant_age	Incomplete data: bij 38,5% van de incidenten is deze kolom leeg.
participant_age_group	Te breed: er wordt enkel onderscheid gemaakt tussen kinderen (0-11), tieners (12-17) en volwassenen (18+).
participant_gender	Niet relevant voor onderzoeksvragen
participant_name	Niet relevant voor onderzoeksvragen
participant_relationship	Incomplete data: bij 93,4% van de incidenten is deze kolom leeg.
participant_status	Overbodig: relevante waarden kan uit andere kolommen worden verkregen.
participant_type	Niet relevant voor onderzoeksvraag
sources	Het is niet van belang ieder incident individueel te bekijken.
state.house_district	Per incident is informatie over de staat en stad beschikbaar, incidenten afbakenen op district is overbodig.
state.senate_district	Per incident is informatie over de staat en stad beschikbaar, incidenten afbakenen op district is overbodig.

Table 2: Dataset *Gun Violence*: wapengerelateerde incidenten

Kolom	Type	Omschrijving	Attribuut-type
incident_id	int	<i>gunviolencearchive.org</i> ID voor incident	nominaal
date	str	datum van incident	ordinaal
state	str	staat	nominaal
city_or_county	str	stad of provincie	nominaal
n_killed	int	aantal mensen gedood	ratio
n_injured	int	aantal mensen gewond	ratio
incident_characteristics	list[str]	lijst van kenmerken incident	nominaal

Table 3: Dataset *Political Affiliation*: politieke voorkeur

Kolom	Type	Omschrijving	Attribuut-type
state	str	naam van staat, inclusief <i>District of Columbia</i>	nominaal
democrat	int	percentage Democratische stemmen in verkiezingen 2016	ratio
republican	int	percentage Republikeinse stemmen in verkiezingen 2016	ratio
other	int	percentage stemmen niet Democratisch en niet Republikeins in verkiezingen 2016	ratio



Table 4: Dataset *Gun Ownership*: wapenbezit

Kolom	Type	Omschrijving	Attribuut- type
state	str	naam van staat, inclusief <i>District of Columbia</i>	nominaal
population	int	aantal inwoners van de staat	ratio
guns_registered	int	aantal geregistreerde wapens in de staat	ratio
guns_per_capita	int	gemiddeld aantal wapens per persoon in de staat	ratio

Table 5: Gebruikte modules

Software	Versienummer
Python	3.6.4
Jupyter Notebook	5.4.0
Numpy	1.14.0
Pandas	0.22.0
Bokeh	0.12.13
Scikit-learn	0.19.1