

# Vuurwapengeweld in de USA

---

28 juni 2018

*Student:*

Anna Stalknecht, Bart van Laatum,  
Lisa Pleijt, Melanie Baaten  
10792872, 10812431, 11013125,  
11053909

*Tutor:*

Nick de Wolf

*Cursus:*

Data Analysis and Visualization

## 1 Introductie

"The U.S. Is in a Different World" kopt de New York Times [Kevin Quealy, 2016] na de mass shooting in Las Vegas in 2016. Naar aanleiding van deze shooting zijn ze het aantal slachtoffers door vuurwapens gaan vergelijken met andere welvarende landen. De New York Times vergelijkt het aantal doden door vuurwapens per dag met andere rijke landen waarbij genormaliseerd wordt op het inwoneraantal. De VS scoort wel vijf keer hoger dan zijn opvolger. Voor het vak Data Analysis and Visualization is er data over geweld met vuurwapens in de VS in handen gekregen en is hier een onderzoek mee gedaan. Welke interessante patronen worden er gevonden in deze data? Dit is onderzocht aan de hand van de drie gegeven hoofdvragen:

1. Zijn er opmerkelijke verschillen tussen staten? En tussen jaren?
2. Bevat de dataset patronen die je niet zou verwachten?
3. Welke patronen tussen slachtoffer en dader kun je vinden over de jaren ten opzichte van het type incident?

Behalve de al gegeven hoofdvragen is er ook gezocht naar andere interessante verbanden. Zo blijkt uit onderzoek dat agressief gedrag toeneemt naarmate de temperatuur in de leefomgeving hoger wordt [A.Anderson, 2000]. In dit onderzoek wordt er onderzocht of dit in de gegeven data ook het geval is door te kijken naar een verschil in aantal incidenten in de warme en de koude maanden.

Na het onderzoeken van de dataset bleek dat dit onderzoek het beste verdeeld kon worden in drie thema's: geslacht, locatie en categorieën. Tussen deze thema's bevindt zich uiteraard een overlap. In de methode zal aan de hand van deze thema's de werkwijze worden uitgelegd. De drie hoofdvragen zullen ook worden beantwoord in het licht van deze thema's. In het thema "geslacht" zal worden gekeken naar de verhouding in de victim/suspect relatie over de jaren, leeftijden en categorieën. Het thema "locatie" behandelt de vraag: wat voor soort wapengeweld gebeurt waar? Tot slot zal in het thema "categorieën" zoek worden gegaan naar opvallende verschillen of overeenkomsten tussen incident types en het geslacht, leeftijd en locatie van zowel daders als slachtoffers.

## 2 Methode

### 2.1 Cleaning

Voor dit onderzoek is een dataset in een CSV formaat gebruikt die 29 kolommen bevat. Al deze kolommen zijn terug te vinden onder appendix A. Omdat niet al deze kolommen bruikbaar waren voor onze doeleinden zijn er een aantal verwijderd. De volgende kolommen zijn verwijderd: `gun_stolen`, `n_guns_involved`, `participant_relationship`, `notes`, `location_description`. Deze kolommen zijn verwijderd omdat ze minder dan 40% van de incidents ingevuld waren. Daarnaast zijn de volgende kolommen ook verwijderd: `Incident_url`, `source_url`, `incident_url_fields_missing`, `sources`. Deze kolommen bevatte informatie van waar de informatie over dat specifieke incident vandaan kwam. Deze kolommen voegde geen waarde toe voor het visualiseren van de data. Van de overige data is met behulp van pandas een json bestand gemaakt. In tabel 1 zijn de kolommen terug te vinden die uiteindelijk gebruikt zijn in de dataset.

<code>incident_id</code>	<code>date</code>	<code>state</code>	<code>city_or_county</code>
<code>address</code>	<code>n_killed</code>	<code>n_injured</code>	<code>congressional_district</code>
<code>gun_type</code>	<code>incident_characteristics</code>	<code>latitude</code>	<code>longitude</code>
<code>participant_age</code>	<code>participant_age_group</code>	<code>participant_geslacht</code>	<code>participant_name</code>
<code>participant_status</code>	<code>participant_type</code>	<code>state_house_district</code>	<code>state_senate_district</code>

Tabel 1: Gebruikte kolommen

Er is gekozen voor json omdat in de dataset bij bepaalde velden meerdere waarden en meerdere waarden met 'key' vermeld stonden. In json konden deze waarden gemakkelijk gegroepeerd worden als lijst of als nieuwe dict om zo de relatie tussen deze waardes te behouden. Om de datum gemakkelijker te gebruiken is besloten ook deze in een lijst te zetten: `[yyyy,mm,dd]`.

### 2.2 Geslacht

In het thema geslacht lag de focus op de volgende vraag: Hoe verhoudt de victim/suspect relatie zich over de jaren, leeftijden en categoriën. Er zijn plots gegenereerd die een kwalitatieve vergelijking maken tussen victims en suspects van vrouwen en mannen en hoe vaak mannen en vrouwen voorkomen in gedocumenteerde data. Daarnaast wordt er ook gekeken naar de verschillende incident categorieën en hoe het geslacht zich hier tussen verhoudt. Verder zijn er voor vier jaren plots gemaakt die kwalitatieve vergelijkingen bieden tussen staten en hoe de geslachten, slachtoffers en verdachten zich voor elke staat verhouden.

### 2.3 Kaart

In deze subsectie is er een kaart van de USA geplot, deze kaart is helaas exculsief de staten Alaska en Hawaïi. Er is besloten om deze weg te laten aangezien deze staten slechts een zeer klein aandeel van het aantal incidenten bevat. Er is gekozen voor een kaart visualisatie omdat deze kan laten zien of er clustering plaatsvindt. Verschillende data zal op een interactieve wijze op deze map worden geplot. Zo zijn alle incidenten terug te vinden op de kaart, is er een heatmap van de staten gemaakt en zijn ook de categorieën er terug te vinden. Dit zodat er te zien is waar er meerdere incidenten zijn en om zo ook een kwalitatieve vergelijking te maken tussen staten en specifieke categorieën. Zijn bepaalde categorieën vaker terug te vinden in een specifieke staat? Verder is er een heatmap van de incidenten per staat gemaakt.

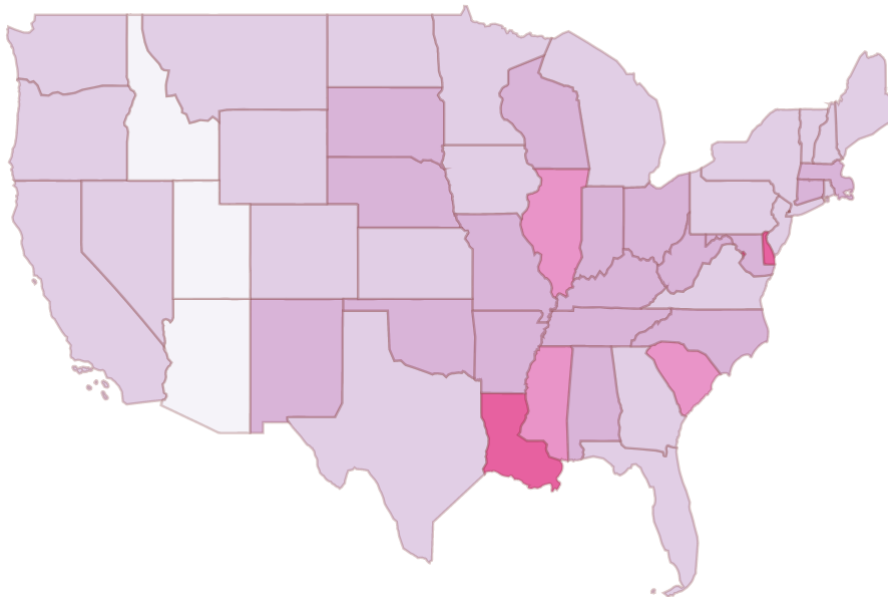
## 2.4 Categorieën

In de dataset werden er categorieën toegewezen aan de incidents in de kolom 'incident\_type'. Een incident kon meerdere categorieën hebben, slechts een enkele keer waren er geen categorieën toegewezen aan een incident. Deze incidenten zijn tijdens het onderzoek naar incident types niet meegeteld. In totaal waren er 109 categorieën die aan een incident konden worden meegegeven. Omdat er geconstateerd werd dat een groot aantal categorieën in grote lijnen onder eenzelfde overkoepelende categorie vielen is de keuze gemaakt om de categorieën te reduceren naar 22 categorieën. Hierbij zijn ook een aantal categorieën buiten beschouwing gelaten omdat deze zeer breed waren en bij vrijwel elk incident aanwezig waren. Ook zijn een aantal categorieën verwijderd omdat ze te weinig werden gebruikt, of pas sinds 2017 geïntroduceerd zijn. Zowel de oude als de nieuwe verdeling van de categorieën is terug te vinden onder appendix B. Hierna is een plot gegenereert die een kwalitatieve vergelijking maakt tussen de suspects en victims per jaar, per geslacht, per age, per state en per categorie.

## 3 Resultaten

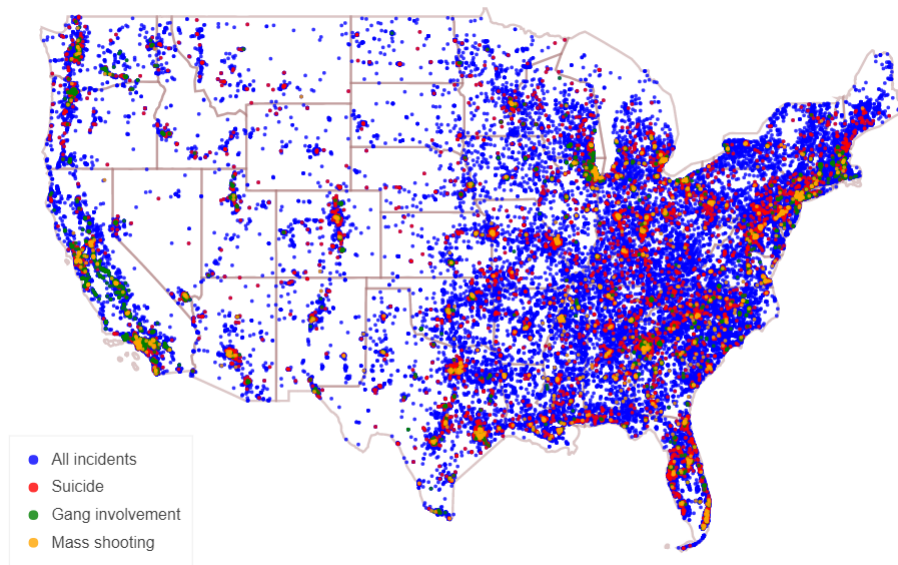
### 3.1 Zijn er opmerkelijke verschillen tussen staten? En tussen jaren?

In figuur 1 is een heatmap van Amerika te zien waarin de verschillen in het aantal incidenten tussen de staten goed duidelijk wordt. Hierbij is ook rekening gehouden met de bevolkingsdichtheid van de staten. Uitschieters zijn Louisiana (onderin op de kaart) en Delaware (de kleine staat rechts).

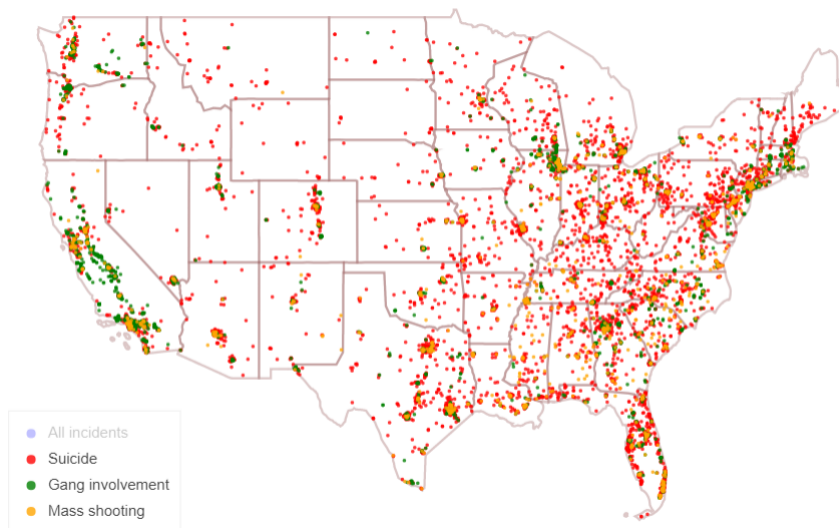


Figuur 1: Heatmap staten van de US

Naar aanleiding van de heatmap zijn we de incidenten als puntjes op de kaart gaan plotten, waarbij we een aantal categorieën hebben uitgekozen om de verschillen tussen de staten duidelijk te maken. Deze is te zien in figuur 2. Een interactieve versie van deze kaart is te vinden in de github in de docs folder. Bij het uitvinken van de categorie "all incidents" zijn de verschillen tussen de categorieën goed te zien. Zo is "suicide" een incident wat redelijk verspreid over de kaart gebeurt. Terwijl "gang involvement" meer voorkomt in de oost- en westkust en in de steden. Zie hiervoor figuur 3.



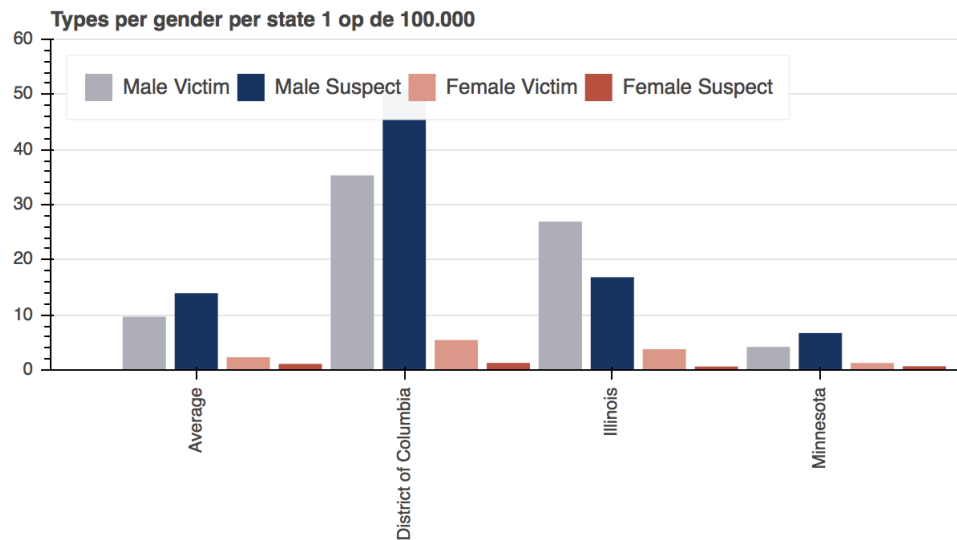
Figuur 2: Incidenten per categorie op locatie



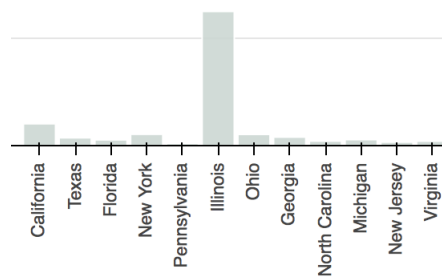
Figuur 3: Suicide, gang involvement en mass shooting

Verder is er ook gekeken naar verschillen in type participant per staat. Het viel op dat de resultaten hier best wel uit elkaar kunnen lopen. Hier onder zijn een aantal opmerkelijke resultaten weergegeven. De volledige plot is terug te vinden in de github in de plot folder onder de naam `geslacht_states_indicents`. In een aantal staten was er te zien dat er meer victims vielen dan er suspects waren. Een voorbeeld hiervan is Illinois. Hier zien we over het algemeen meer gunviolence. Veel hiervan is drugrelated gunviolence en gang-shootings (zie figuur 6). Aangenomen wordt dat er bij dit soort gunviolence niet vaak een suspect wordt gevonden. Verder viel het op dat in district of Colombia veel meer gunviolence plaatsvindt dan in andere states. Er wordt gedacht dat dit komt omdat het een stadsdeel is en er over het algemeen een hogere bevolkingsdichtheid heeft. Wat verder nog opvallend was waren staten waarin juist relatief best weinig gunviolence is, zoals Minnesota.

Over de jaren genomen zijn zowel het aantal gewonden door vuurwapengeweld als het



Figuur 4: Type participant voor enkele staten per geslacht. 1 op de 100.000



Figuur 5: Kwalitatieve vergelijking van staten voor gangrelated shootings

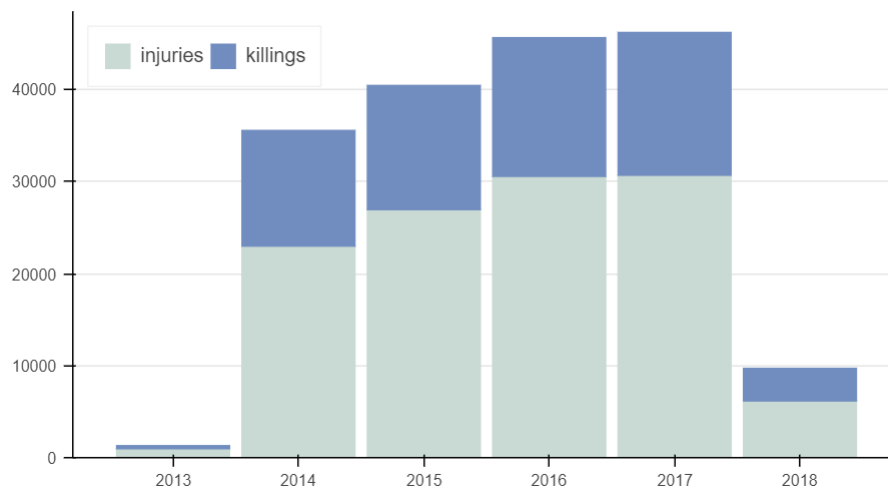
aantal doden toegenomen.

### 3.2 Bevat de dataset patronen die je niet zou verwachten?

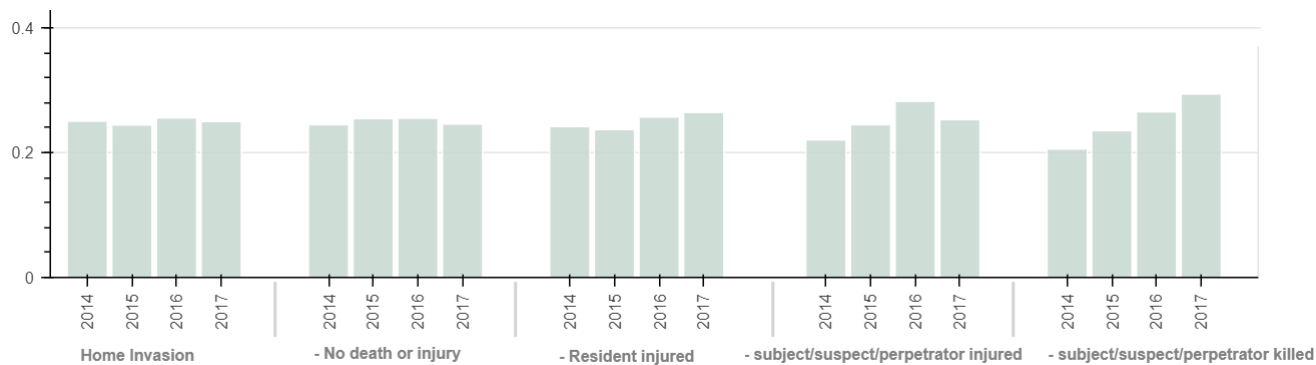
Tijdens het plotten zijn er zeker patronen tegengekomen die we niet zouden verwachten. Voornamelijk kijkend naar categorieën zijn er een aantal dingen onverwacht. Zo viel op dat in de categorie home invasion het aantal home invasions over de jaren ongeveer gelijk blijft maar het aantal victims wat bij een home invasion wordt vermoord toeneemt. De home invasions worden dus steeds agresiever.

Bij het kijken naar het aantal suicides per maand kwamen we ook een onverwacht patroon tegen. Bij het lezen van een artikel over maanden waarin suicides meer voorkwamen [sui, 2013]. Bleek namelijk dat in december het suicide aantal het laagst is. Zelfmoord komt het meeste voor in de lente en de herfst. Figuur 5 laat zien hoe vaak suicide per maand in onze data voorkomt. In onze data is inderdaad de dip in december te zien, dit is dus naar verwachting van het artikel. Maar de piek in de lente en de herfst is bij ons niet van toepassing. In onze data piekt het zelfmoordaantal vooral in de zomer.

Daarnaast zien we over alle jaren duidelijk een dip in het aantal incidenten in februari. Deze dip kon eigenlijk niet echt verklaard worden. Als we de categorieën per maand plotten kwamen hier geen grote verschillen in februari uit namelijk.



Figuur 6: Home invasion



Figuur 7: Home invasion

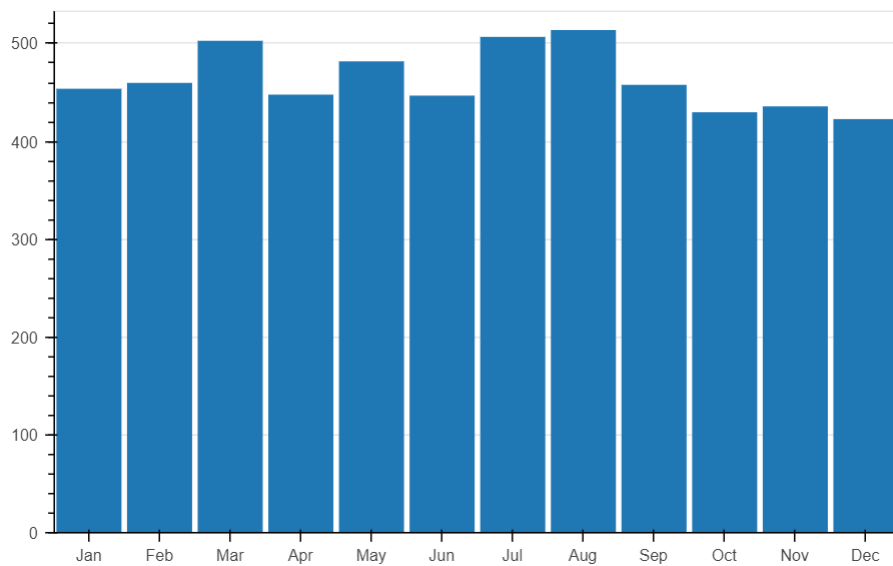
### 3.3 Wat zijn de patronen tussen slachtoffer en dader ten opzichte van type incident?

Een patroon in de type incidenten is voornamelijk te vinden in de leeftijden van zowel de slachtoffers als de daders. Deze is vaak ongeveer hetzelfde. Zie onderstaande figuren waar een aantal categorieën zijn weergegeven.

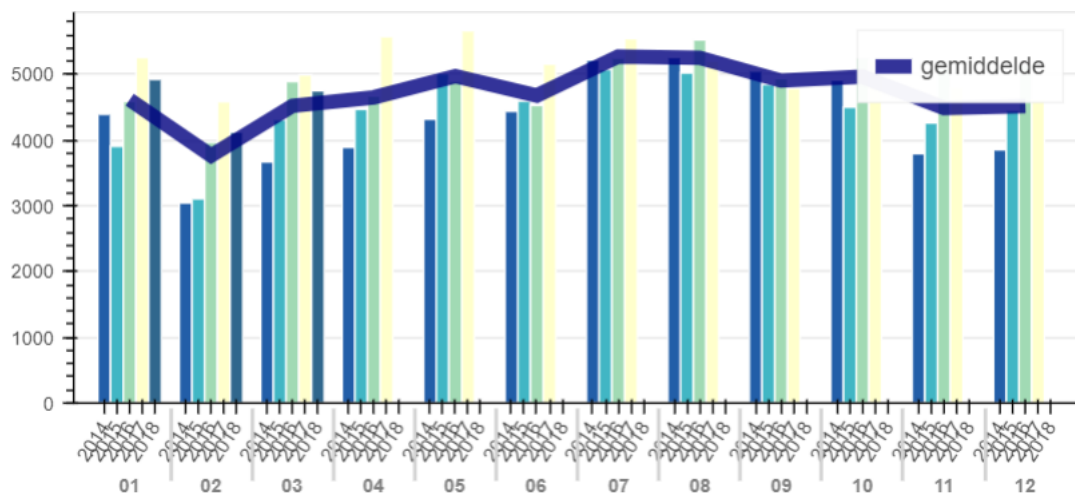
### 3.4 Is er meer agressie in de zomermaanden?

Om uit te zoeken of het onderzoek over meer agressie bij warmte ook uit onze dataset blijkt is er eerst een plot gemaakt met het aantal incidenten over de jaren per maand. Om te kijken of er een piek is in de warme maanden.

Daarna is er ook per categorie hiernaar gekeken en ook hieruit bleek dat de agressievere categorieën een grotere kans hadden op meer incidenten in de warme maanden.



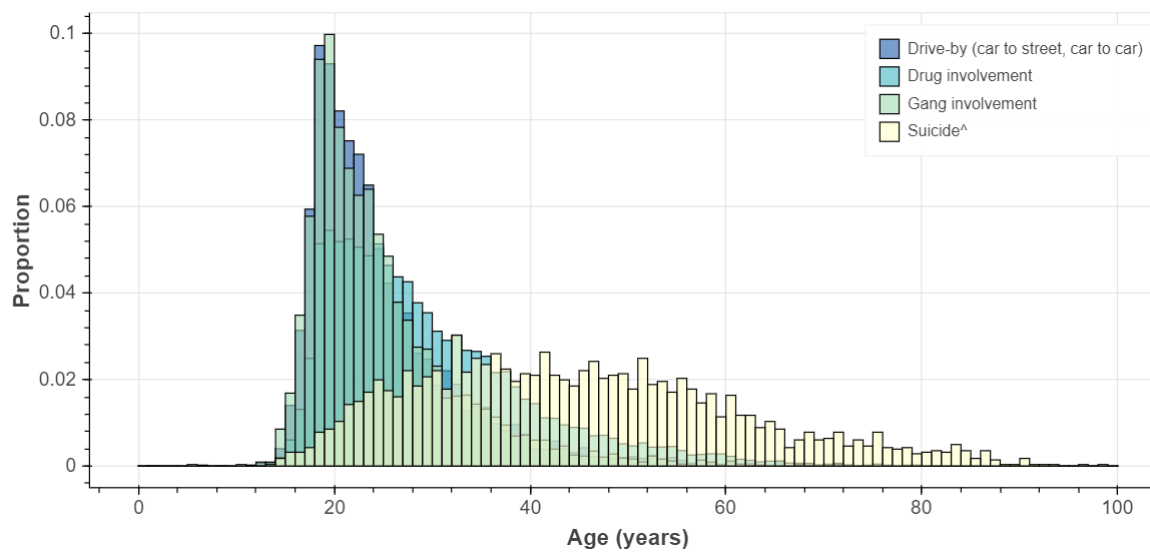
Figuur 8: Aantal keer zelfmoord per maand (geteld over 4 jaar)



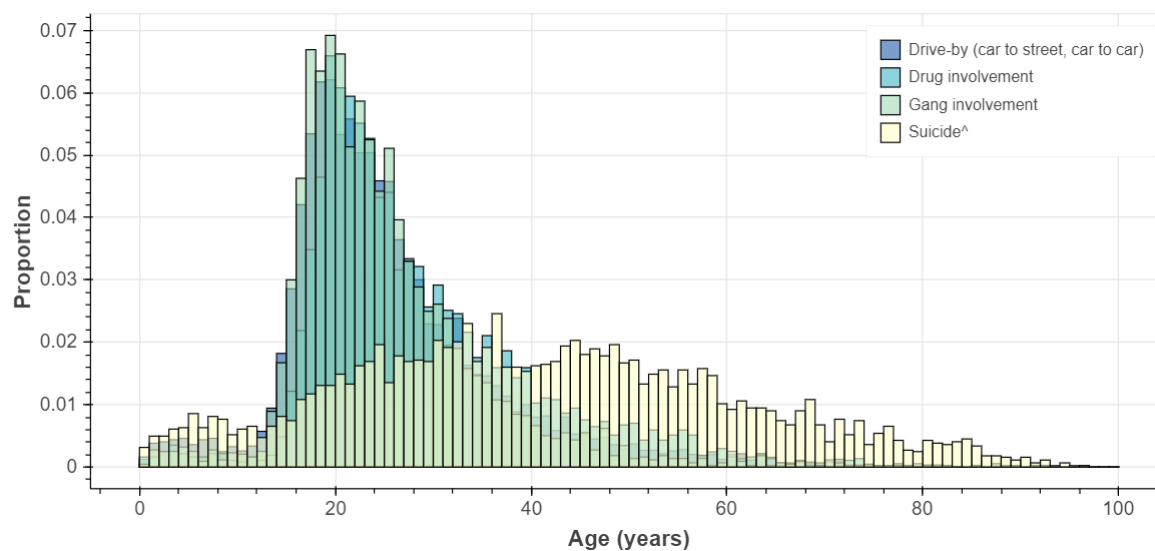
Figuur 9: Incidenten per maand per jaar

## 4 Discussie

Dit onderzoek is opgezet aan de hand van drie thema's. Aan de hand van deze thema's zijn de onderzoeksvragen beantwoord. Bij de kaart van Amerika is er gekeken naar de locatie van de incidenten. Er zijn tussen staten een aantal verschillen, zo is in de staat illinois veel meer vuurwapengeweld dan in andere staten. Dit komt waarschijnlijk doordat in Chicago veel vuurwapengeweld is [Ruby, 2016] en deze stad in Illinois ligt. De categorieën 'gang violence' en 'drug involvement' komen voornamelijk voor in de steden en aan de oost- en westkust. Terwijl de categorie 'suicide' veel meer verspreid over de kaart plaatsvindt. Grote verschillen in geslacht zijn vooral te zien in het aantal incidenten. Mannen zijn zowel als victim en als subject meer aanwezig. Wel is het zo dat vrouwen vaker victim zijn dan subject, terwijl dit bij mannen andersom is. Tijdens het plotten van de categorieën bleek dat de meeste



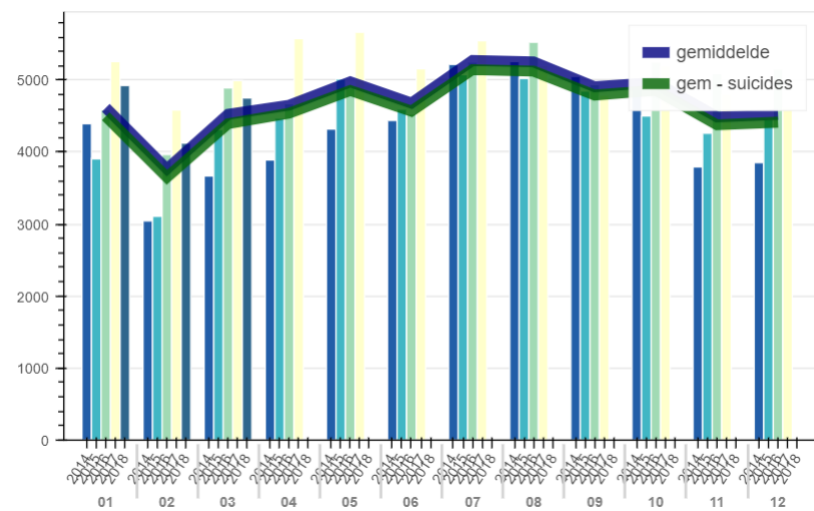
Figuur 10: Aandeel verdachten per leeftijd



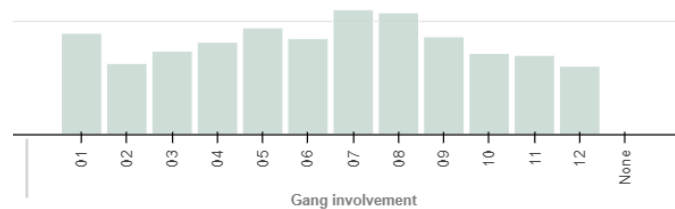
Figuur 11: Aandeel slachtoffers per leeftijd

categorieën veel overeenkomsten hadden in de leeftijd van de subjects, voornamelijk rond de 20 zat hierin een piek, enkele categorieën waren hierin afwijkend. In vervolg onderzoek is het voornamelijk interessant te kijken naar de dip van aantal incidenten in februari die in alle jaren voorkomt. Ook liepen wij bij vergelijkingen over de jaren er tegenaan dat we van 2018 slechts drie maanden aan data hadden. Hierdoor kon er moeilijk worden gezien of een stijgende of zakkende trend door zou zetten of dat dit slechts over een jaar zo was. Met meer (consistent verzamelde) data en meer jaren, kan hier veel meer over gezegd worden.

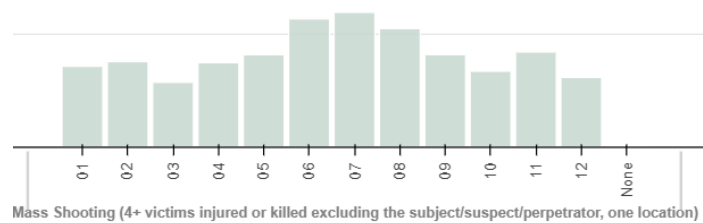




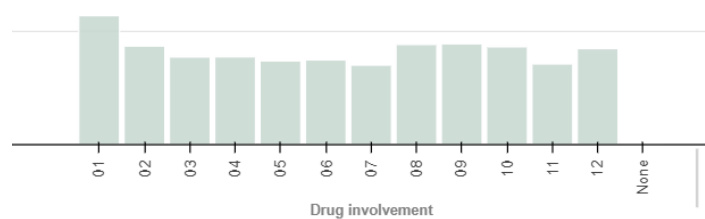
Figuur 12: Aantal incidenten per maand per jaar



Figuur 13: Gangviolence over de maanden



Figuur 14: Mass shootings over de maanden



Figuur 15: Drug involvement over de maanden

## Referenties

Margot Sanger-Katz Kevin Quealy. Comparing gun deaths by country: The u.s. is in a different world. *New York Times*, 2016.

Craig A.Anderson. Temperature and aggression. *Elsevier*, 2000.

Holiday suicides: Fact or myth? *Center for disease control and prevention*, 2013.

Matt Ruby. A weekend in chicago. *New York Times*, 2016.

## 5 Appendix A

Alle kolommen:

incident\_id  
date  
state  
city\_or\_county  
address  
n\_killed  
n\_injured  
incident\_url  
source\_url  
incident\_url\_fields\_missing  
congressional\_district  
gun\_stolen  
gun\_type  
incident\_characteristics  
latitude  
location\_description  
longitude  
n\_guns\_involved  
notes  
participant\_age  
participant\_age\_group  
participant\_geslacht  
participant\_name  
participant\_relationship  
participant\_status  
participant\_type  
sources  
state\_house\_district  
state\_senate\_district

## 6 Appendix B

Op de volgende pagina is terug te vinden hoe de categorieën tot stand zijn gekomen. Alles wat een tabje verder staat onder de bovenste regel zijn oude categorieën die vanaf nu mee worden genomen in de bovenstaand overkoepelende categorie.

## Raids

- ATF/LE Confiscation/Raid/Arrest
- Stolen/Illegally owned gun{s} recovered during arrest/warrant
- Possession (gun(s) found during commission of other crimes)

## Accidental Shooting

- Accidental Shooting - Death
- Accidental Shooting - Injury
- Accidental Shooting at a Business
- Accidental/Negligent Discharge
- Cleaning gun
- Playing with gun
- Thought gun was unloaded
- Self-Inflicted (not suicide or suicide attempt - NO PERP)

## Bar/club incident - in or around establishment

## Child Involved Incident

- Child injured (not child shooter)
- Child injured by child
- Child injured self
- Child killed (not child shooter)
- Child killed by child
- Child killed self
- Child picked up & fired gun
- Child with gun - no shots fired

## Defensive Use

- Defensive Use - Crime occurs, victim shoots

## subject/suspect/perpetrator

- Defensive Use - Good Samaritan/Third Party
- Defensive Use - Shots fired, no injury/death
- Defensive Use - Stand Your Ground/Castle Doctrine established
- Defensive Use - Victim stops crime
- Defensive Use - WITHOUT a gun
- Defensive use - No shots fired

## Domestic Violence

- Drive-by (car to street, car to car)

## Drug involvement

## Gang involvement

## Road rage

## Gun at school

- Gun at school, no death/injury - elementary/secondary school
- Gun at school, no death/injury - university/college

## Hate crime

## Home Invasion

- Home Invasion - No death or injury
- Home Invasion - Resident injured
- Home Invasion - Resident killed
- Home Invasion - subject/suspect/perpetrator injured
- Home Invasion - subject/suspect/perpetrator killed

## Institution/Group/Business

- Workplace shooting (disgruntled employee)

## Kidnapping/abductions/hostage

## Mass Problem

- Mass Murder (4+ deceased victims excluding the subject/suspect/perpetrator, one location)

- Mass Shooting (4+ victims injured or killed excluding the subject/suspect/perpetrator, one location)

## Mistaken ID (thought it was an intruder/threat, was friend/family)

## Murder/Suicide

## Officer Involved Incident