

Process

De keuze voor het vak Data Visualisation viel op de case van gunviolence in de US. Dit omdat alle teamleden al snel een idee hadden van wat er interessant zou zijn om te visualiseren. Er is in het begin gekozen om de data in een JSON-file te zetten. Dit omdat er per incident vaak meerdere participanten waren gedocumenteerd. In een JSON-file kon dit makkelijk verwerkt worden in een nested dictionary waar alle participanten opnieuw weer een key worden.

De volgende stap in het process was beslissen welke kolommen we zouden gaan gebruiken. Er is voor gekozen om de kolommen met minder dan 40% aan data weg te laten aangezien er weinig gezegd kon worden over deze kolommen. Dit waren: `gun_stolen`, `n_guns_involved`, `participant_relationship`, `location_description`. Vervolgens waren er een aantal kolommen verwijderd die informatie bevat die niet gevisualiseerd kan worden. Dit waren: `notes`, `incident_url`, `source_url`, `incident_url_fields_missing`, `sources`.

Hierna konden we beginnen met het kijken naar wat voor data we hadden en welke kant we op konden gaan hiermee.

Er werd als eerste besloten om een kaart te plotten met alle incidenten erop. Hiermee konden we namelijk zien of er sprake is van clustering. Ook zou dit ons kunnen helpen om de hoofdvraag op te lossen: 'Zijn er verschillen te zien tussen steden en of staten?'. Om dit uit te breiden is er hierna voor gekozen om ook verschillende categorieën te plotten op deze kaart. Zo konden we ook zien of bepaalde gunviolence crimes meer voorkomen in bepaalde staten. Zo konden we dan ook uiteindelijk iets zeggen over de verschillen die er wel of niet bestaan tussen staten.

In de data waren er maar liefst 109 categorieën te vinden. We zagen dat veel van deze categorieën een overkoepelende categorie hadden. We hebben er daarom voor gekozen om de verschillende categorieën terug te brengen naar 22 categorieën. Ook waren er een aantal categorieën aanwezig die bij vrijwel elk incident aanwezig waren. Er is voor gekozen om deze weg te laten aangezien deze categorie niet veel nieuws zegt over de verschillende incidents.

Verder is ervoor gekozen om te kijken of er verschillen bestaan tussen genders. Hierbij hebben we gekeken tussen de verschillen van participant type. Dus of een vrouw of een man een suspect of victim was. Dit zodat we ook iets konden zeggen over het verschil tussen genders. Er werd vastgesteld dat mannen veel vaker betrokken waren in gunviolence dan vrouwen. En dat ze ook veel vaker een suspect waren. Dit wekte de vraag op of dit in elke staat zo zou zijn. Daarom is er ook gekeken naar de verschillende staten om zo ook een antwoord te kunnen geven op de hoofdvraag "Is er verschil te zien tussen verschillende staten?"

Verder zijn de categorieën ook uitgewerkt en is er gekeken naar bij welke leeftijd welk type gunviolence meer voorkomt. Dit omdat er zo ook iets gezegd kan worden over verschillen tussen leeftijd.

| | | |
|---------|------|---|
| 5 juni | Anna | script gemaakt om csv kolommen te verwijderen en json string te maken |
| 11 juni | Anna | bokeh uitgezocht en bezig aan plot |
| 12 juni | Anna | maand en jaar plot gemaakt |
| 14 juni | Anna | aantal suicides/year/month |
| 15 juni | Anna | piechart guntypes |
| 18 juni | Anna | geplot wanneer de suspect dood gaat |
| 19 juni | Anna | bovenstaande plots mooier gemaakt |
| 20 juni | Anna | eerste concept verslag |

| | | |
|---------|------|--------------------------------|
| 21 juni | Anna | eerste concept verslag |
| 25 juni | Anna | uitzoeken website, verslag etc |
| 26 juni | Anna | website gemaakt |
| 27 juni | Anna | verder met de website |
| 28 juni | Anna | definitief verslag |

| | | |
|---------------|---------|---|
| 5 juni | Melanie | data_types file met welke data typen we hebben en welke data we verwijderen |
| 5 juni | Melanie | converthap.py die de json file op een overzichtelijke manier indent |
| 6 juni | Melanie | nested dictionary gemaakt |
| 7 juni | Melanie | data bekeken |
| 11 juni | Melanie | bokeh bekeken, documentatie eda gelezen en gender.py gemaakt |
| 13 juni | Melanie | gender types op maand plot |
| 15 juni | Melanie | gemiddeld.py gemaakt |
| 16 juni | Melanie | gender states van de jaren gemaakt. en populationstate2017 gemaakt |
| 18 19 20 juni | Melanie | interactief proberen maken van de genderstates |
| 24 juni | Melanie | piechart van genders proberen maken |
| 25 juni | Melanie | states plotten |
| 26 juni | Melanie | Verslag |
| 27 juni | Melanie | code cleanen en gemiddelde proberen maken van alle states |
| 28 juni | Melanie | verslag, process, laatste plots |

| | | |
|---------|------|---|
| 5 juni | Lisa | Deze file gemaakt, git gefixt |
| 6 juni | Lisa | gekeken naar mogelijkheden van de dataset en error gefixt |
| 11 juni | Lisa | bokeh uitgezocht en eerste plots gemaakt |
| 12 juni | Lisa | Dataset verder bruikbaar gemaakt (categorieen indelen) en gekeken naar Kind-gerelateerd wapengeweld |
| 13 juni | Lisa | Verder gekeken naar categorieen |
| 14 juni | Lisa | Slachtoffers en Daders geteld en gekeken of daar iets interessant is, barplots gemaakt |
| 15 juni | Lisa | nog meer barplots |
| 18 juni | Lisa | Ruzie gemaakt met bokeh |
| 19 juni | Lisa | Toch nog meer EDA gedaan, aantallen gewonden of doden per incident |
| 20 juni | Lisa | Script geschreven voor het opslaan van een dict met alle slachtoffers/daders per categorie per jaar/geslacht/leeftijd/staat |
| 21 juni | Lisa | plots gemaakt met data uit bovenstaande dict |
| 22 juni | Lisa | plots gemaakt met data uit dict |
| 25 juni | Lisa | Interactiviteit met bokeh uitzoeken, piecharts zijn blijkbaar onmogelijk |
| 26 juni | Lisa | plot gemaakt (leeftijd/categorieen) in jupyter notebook, lukt niet om deze op github te krijgen |
| 27 juni | Lisa | bestanden uitgezocht en netjes gemaakt, andere plots ook maar in jupyter notebook gemaakt |

| | | |
|---------|------|---|
| 5 juni | Bart | Git geconnect, data gedacht over data visualisaties |
| 6 juni | Bart | Bokeh documentatie bestudeerd, bar chart geplot |
| 11 juni | Bart | bar plot staten gemaakt |
| 11 juni | Bart | map plot documentatie bekeken |
| 12 juni | Bart | Incidents per state per year geplot |