

Terrorisme en het World Happiness Report

In welke mate zijn terroristische aanslag- locaties en tijden voorspelbaar?

Hendrik Huang, Martijn Messanella en Julius Wagt

Inhoudsopgave

- Introductie
 - Achtergrond
 - Dataset(s)
 - Onderzoeksvraag
 - Deelvragen
 - Benodigde Theorie
- Heatmap
- Correlaties tussen datasets
- Fourier-Analyse
- Time Predictor
- Discussie

Achtergrond

- Terrorisme groot en complex probleem
- Gevolgen zijn veel doden, gewonden en schade per jaar
- Prominent agendapunt bij o.a. Tweede Kamerverkiezingen
- Statistische analyse helpt het beter voorspellen of voorkomen van aanslagen
 - Voorbeeld: AIVD kerstquiz, op zoek naar de beste analytici

Datasets

Global Terrorism Database

- Alle terroristische aanslagen tussen januari 1970 en juli 2017
- Meer dan 180.000 aanslagen in totaal
- 135 kolommen, tijd, locatie, letaliteit

World Happiness Index

- Reports van 2016, '17 en '18 meegenomen in de analyse
- Belangrijkste eigenschappen: GDP, vertrouwen in regering,
- Aanname: data in datasets is onafhankelijk van elkaar

Definitie terroristische aanslag

Definition Terrorism according to the GTD:

1. The violent act was aimed at attaining a political, economic, religious, or social goal;
2. The violent act included evidence of an intention to coerce, intimidate, or convey some other message to a larger audience (or audiences) other than the immediate victims; and
3. The violent act was outside the precepts of International Humanitarian Law.

Onderzoeksvraag

- In welke mate zijn terroristische aanslag- locaties en tijden voorspelbaar aan de hand van de GTD en WHI datasets?

Deelvragen

- Is de World Happiness Index (WHI) een voorspeller voor terroristische aanslagen?
- Is er een verband tussen locaties van aanslagen over tijd?
- Zijn er significante frequenties te vinden in de tijden van de aanvallen?
- Volgen de tijden van aanslagen een distributie?
- Kunnen we op locaties een predictor maken voor de tijden?

Theorie

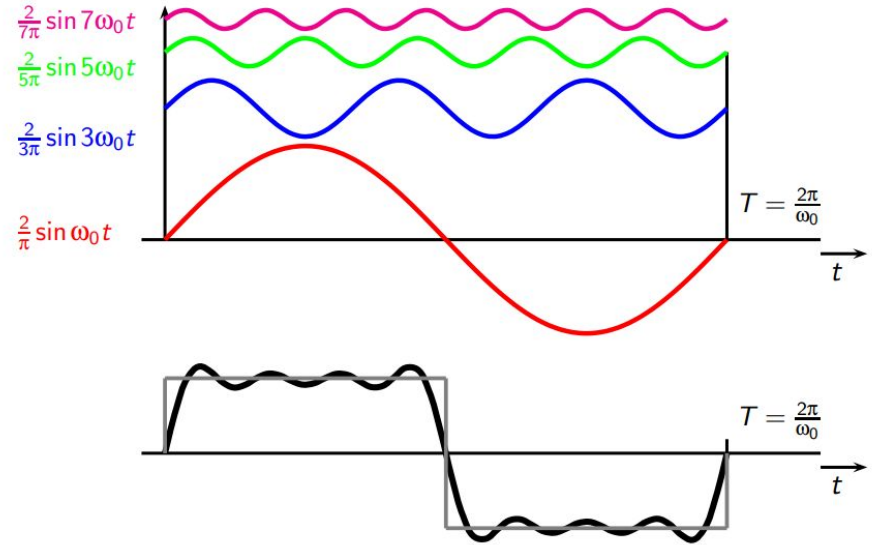
Haversine Distance:

$$\begin{aligned} d &= 2r \arcsin \left(\sqrt{\text{hav}(\varphi_2 - \varphi_1) + \cos(\varphi_1) \cos(\varphi_2) \text{hav}(\lambda_2 - \lambda_1)} \right) \\ &= 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\varphi_2 - \varphi_1}{2} \right) + \cos(\varphi_1) \cos(\varphi_2) \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)} \right) \end{aligned}$$

Theorie

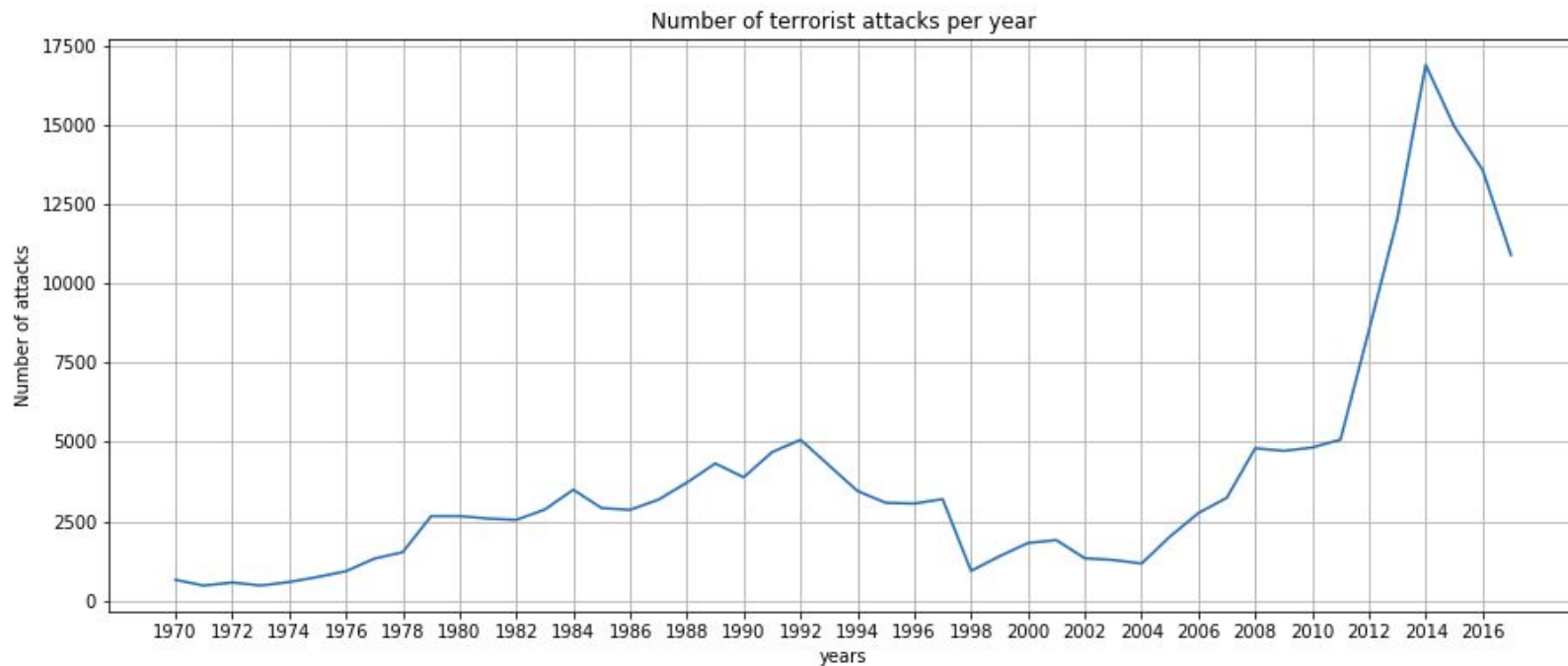
Fourier Transform

- De Fouriertransformatie ontbindt een functie in geschaalde harmonische frequenties.
- Analyse van primaire componenten.

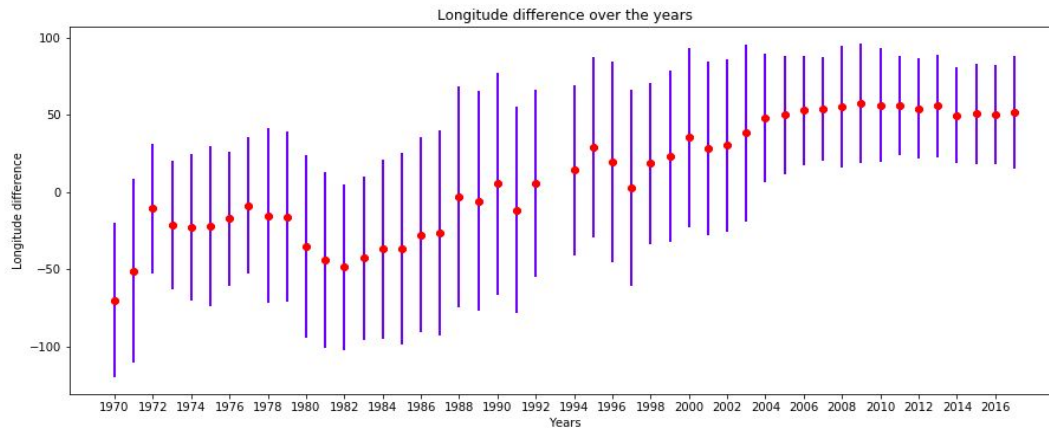
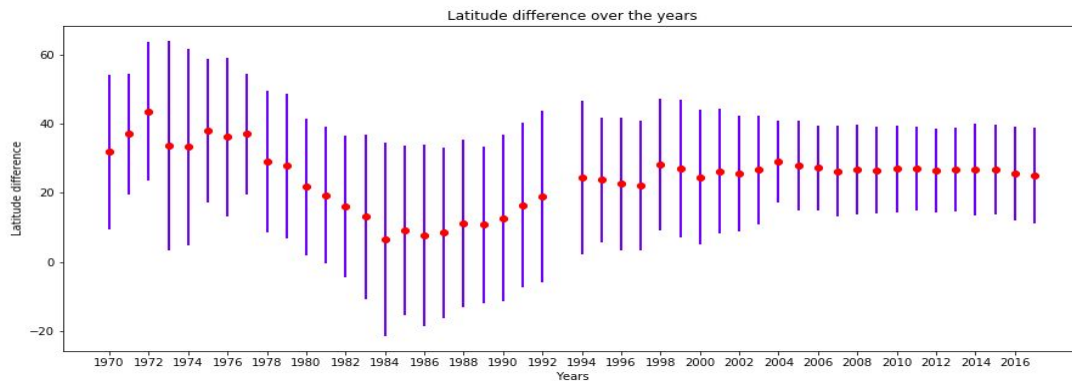


Voorbeeld van een Fouriertransformatie^[1]

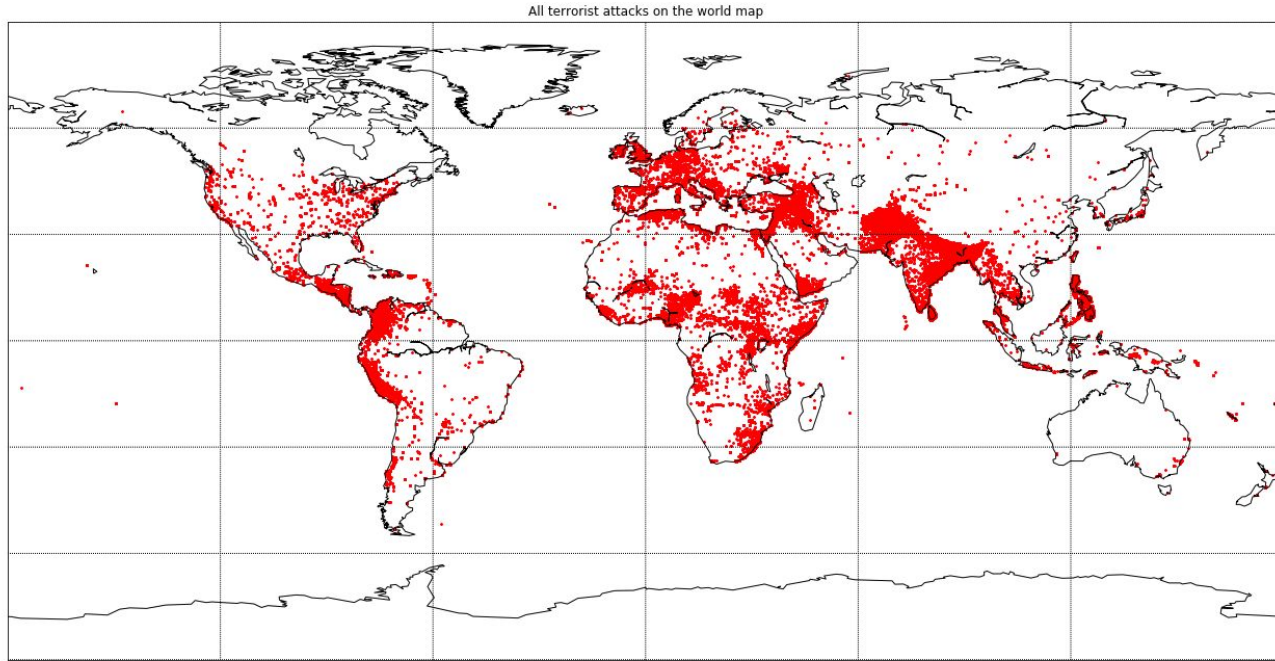
Aanslagen over de jaren



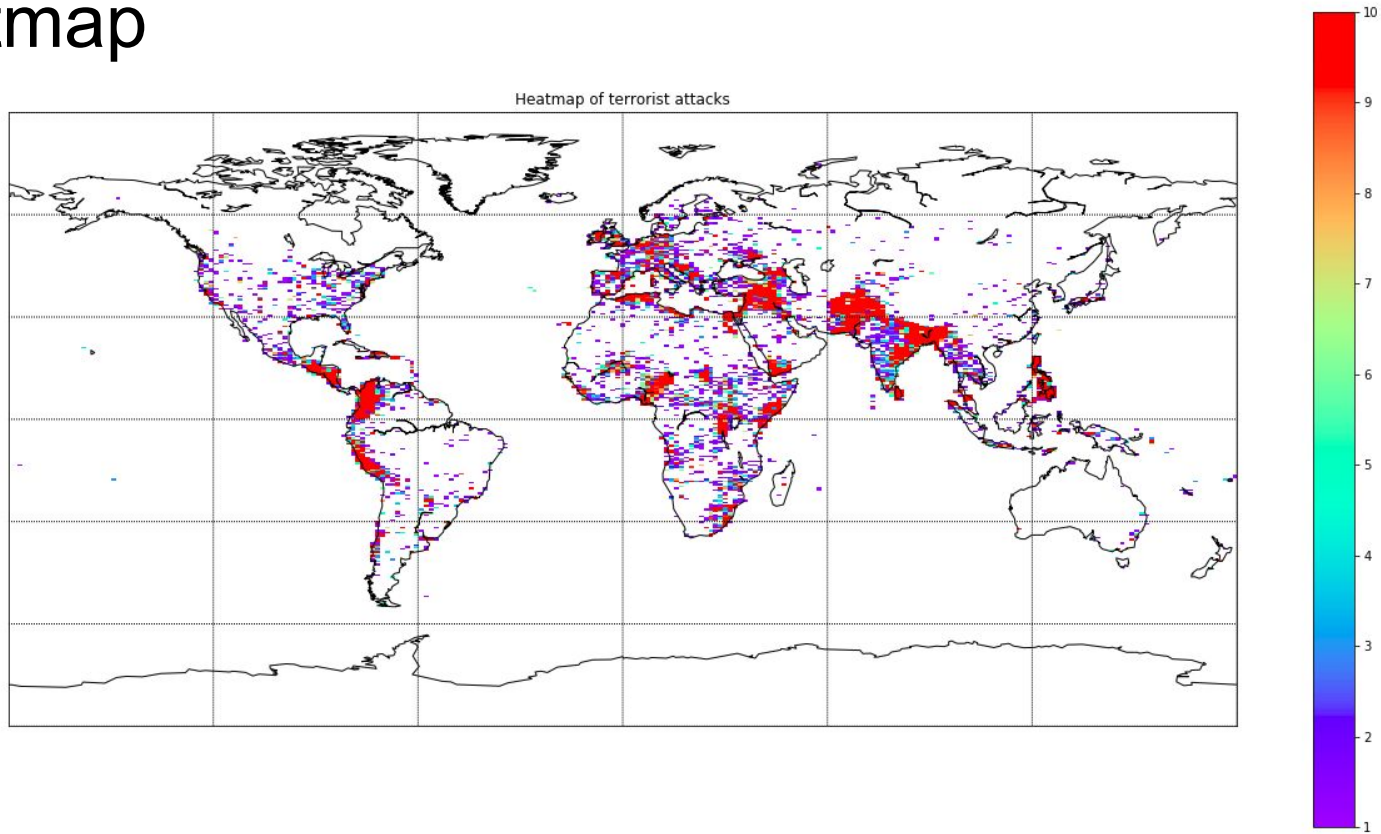
Lengte- en breedtegraden over de jaren



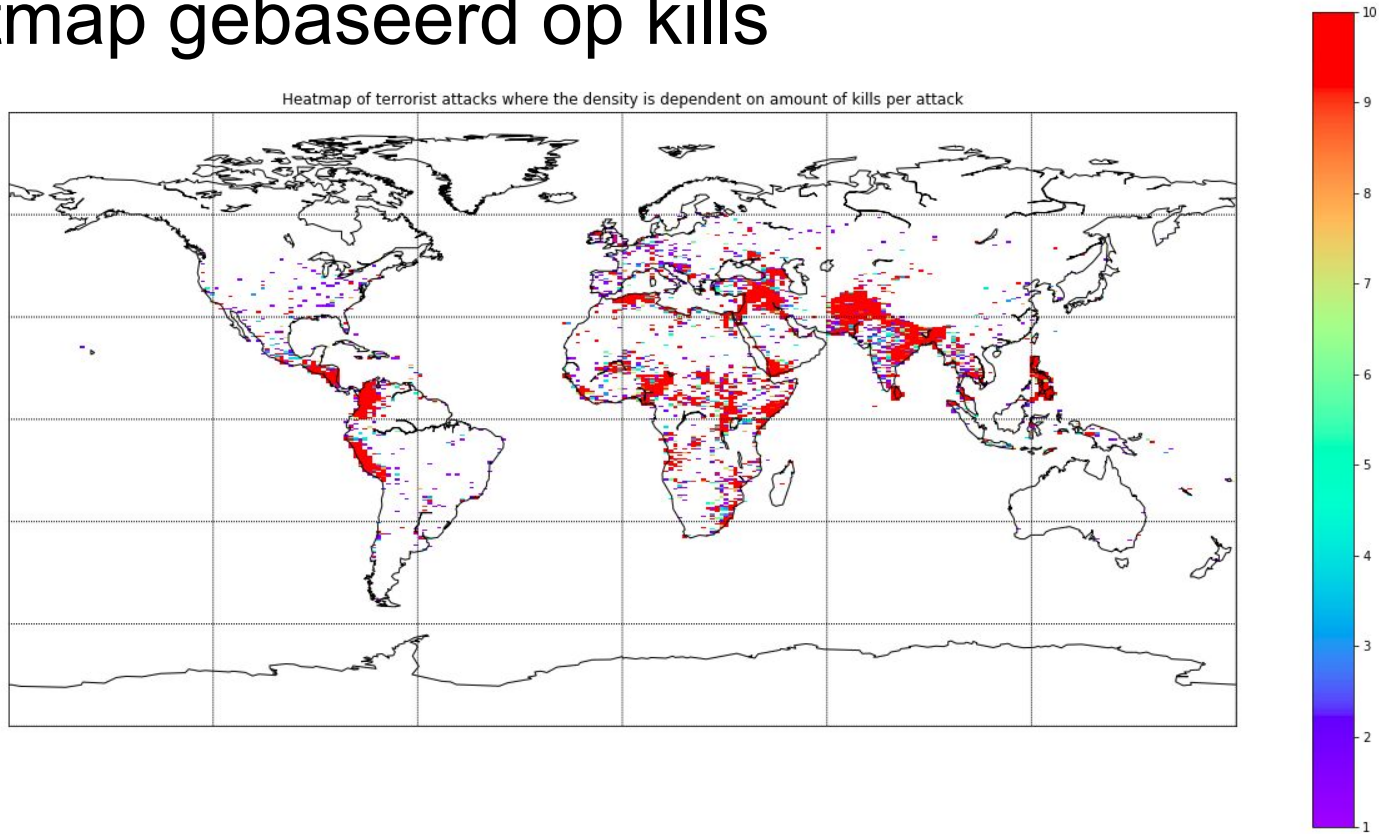
Alle aanvallen geprojecteerd op een wereldkaart



Heatmap

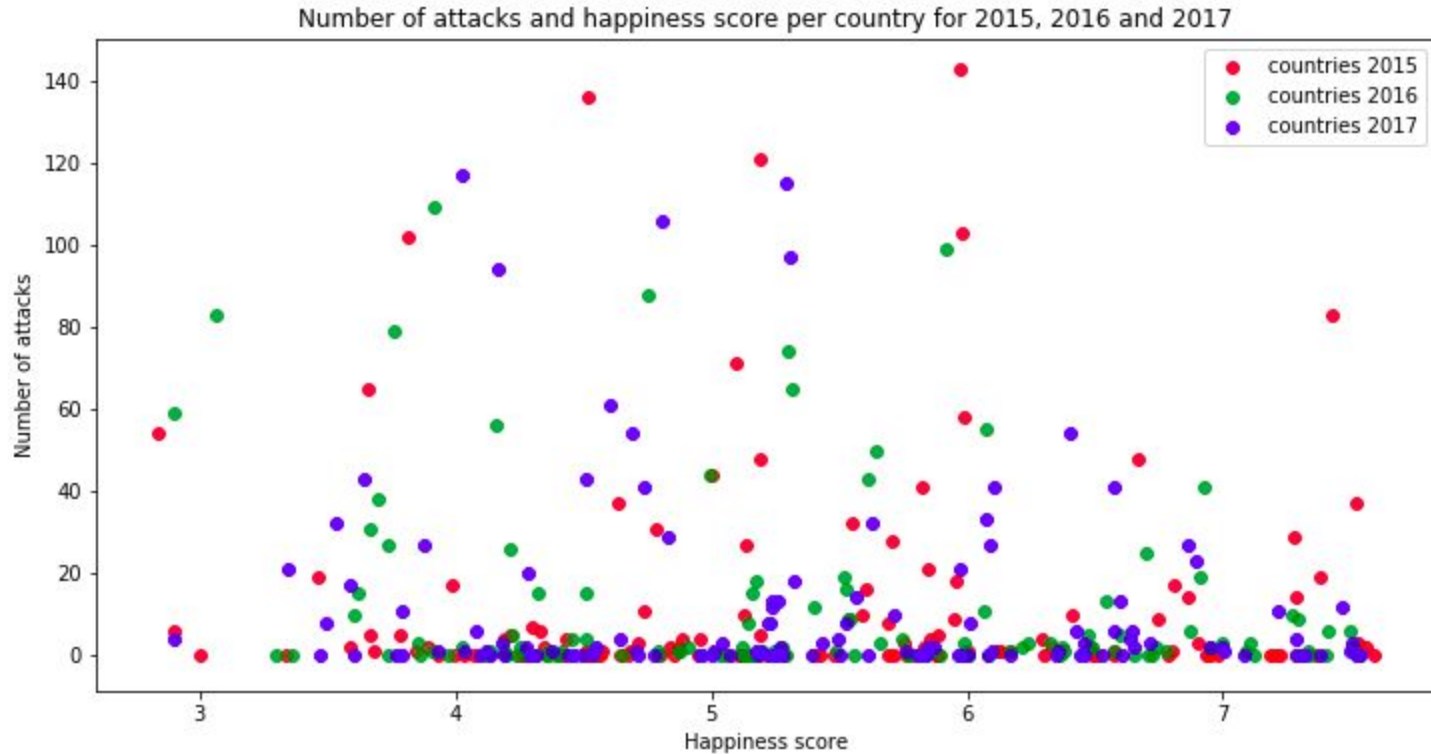


Heatmap gebaseerd op kills

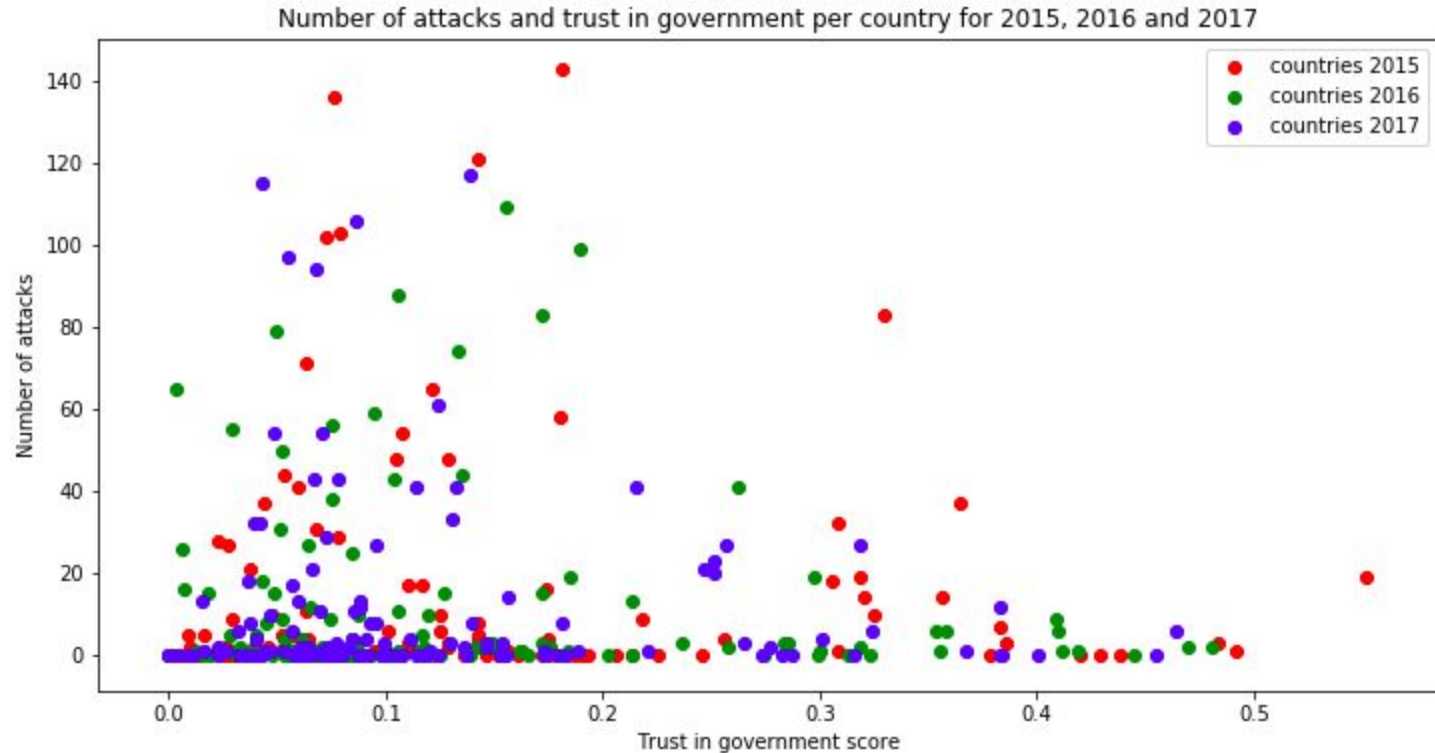


GIF van aanvallen over tijd

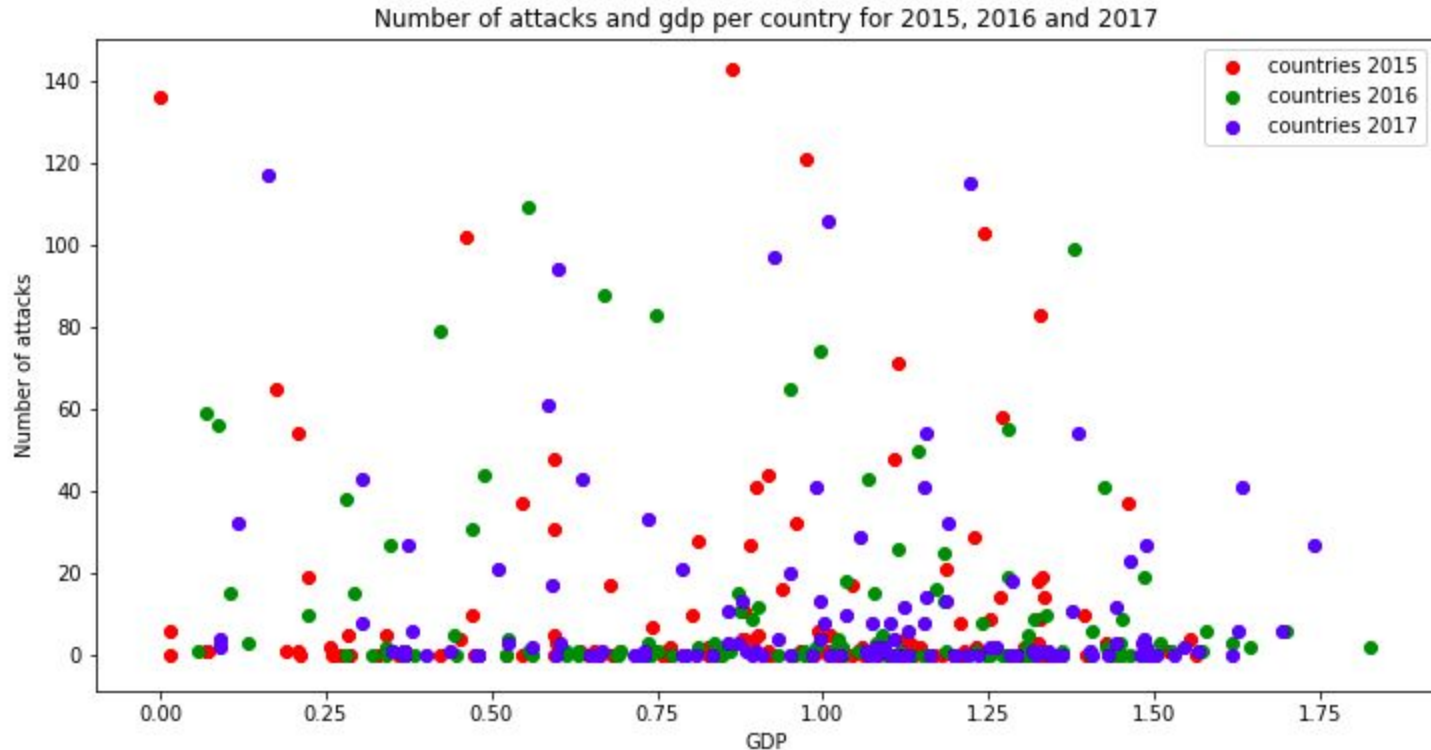
Happiness score tegenover aantal aanvallen per land



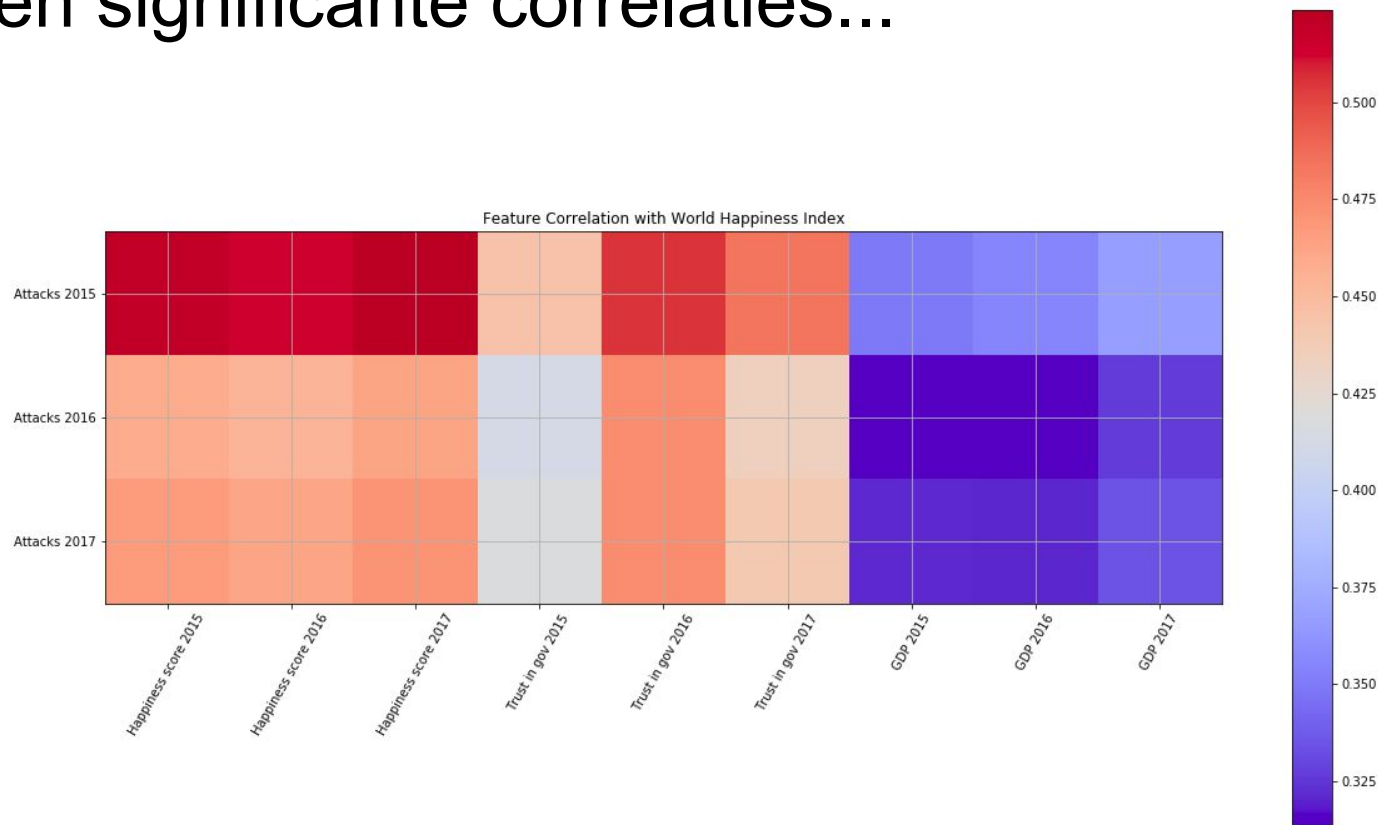
Trust in government tegenover aantal aanvallen per land



GDP tegenover aantal aanvallen per land

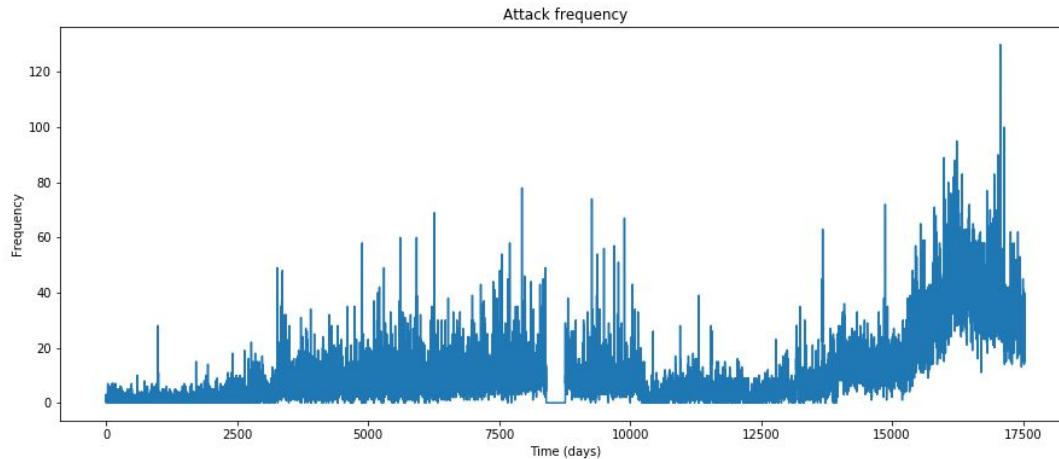


Geen significante correlaties...

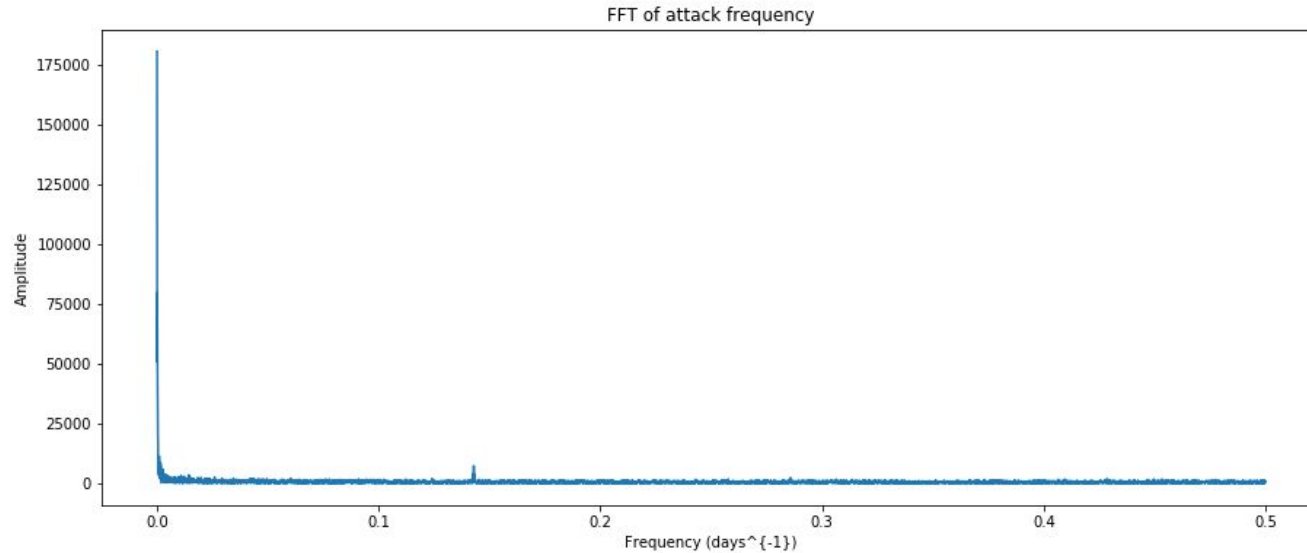


Fourier-Analyse - Methode

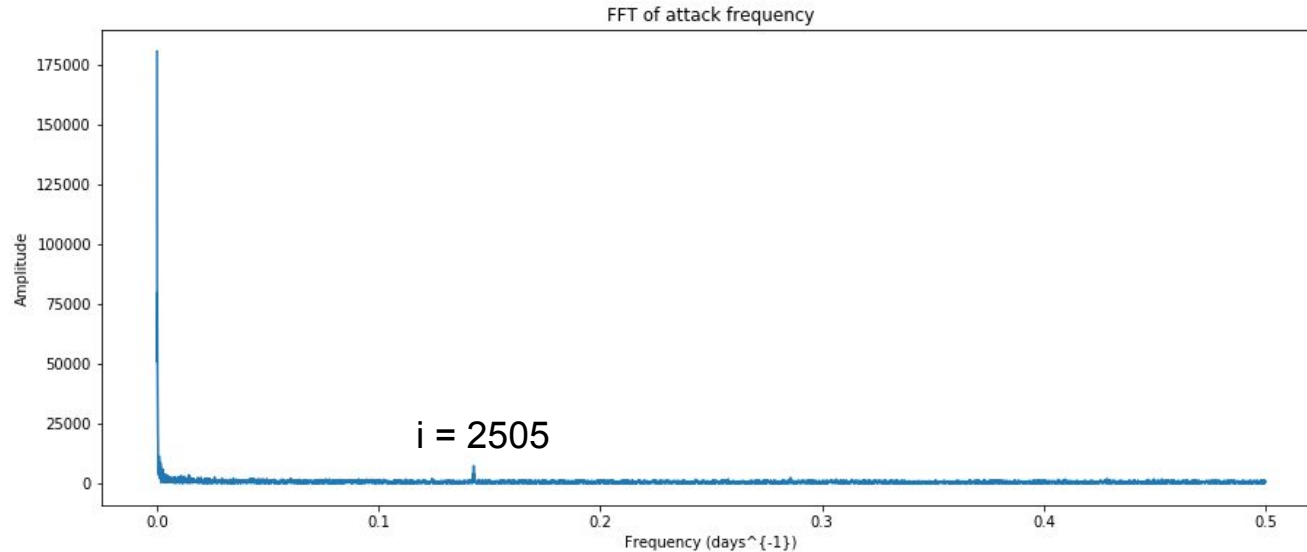
- Transformatie naar amplitude (a) over tijd (t):
 - $t :=$ dagen sinds epoch (1970-1-1)
 - $a :=$ count



Fourier-Analyse - Resultaten

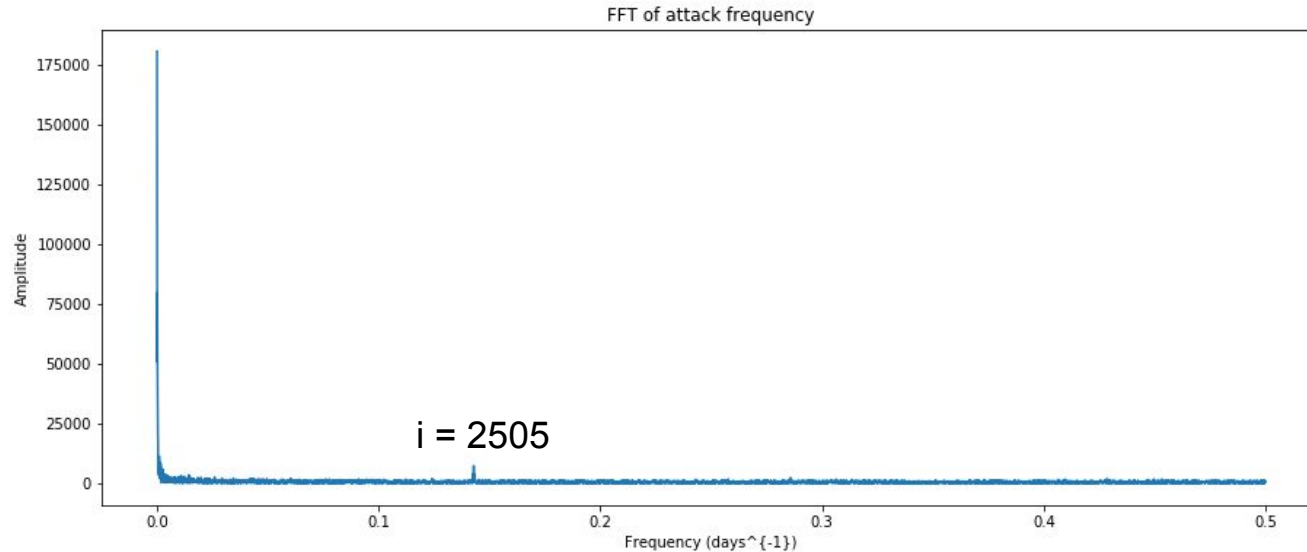


Fourier-Analyse - Resultaten



Piek in frequentie met periode $T = T_0 / \text{index} = ?$

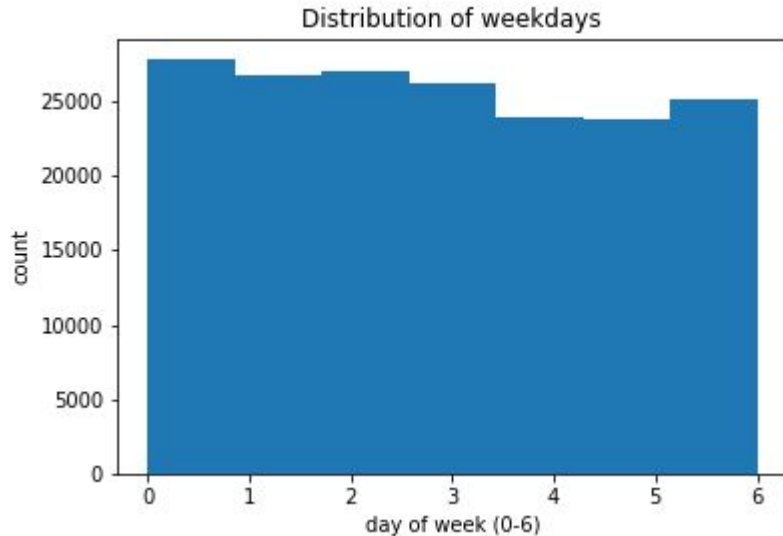
Fourier-Analyse - Resultaten



Piek in frequentie met periode $T = T_0 / \text{index} = 6.998... \approx 7$ dagen

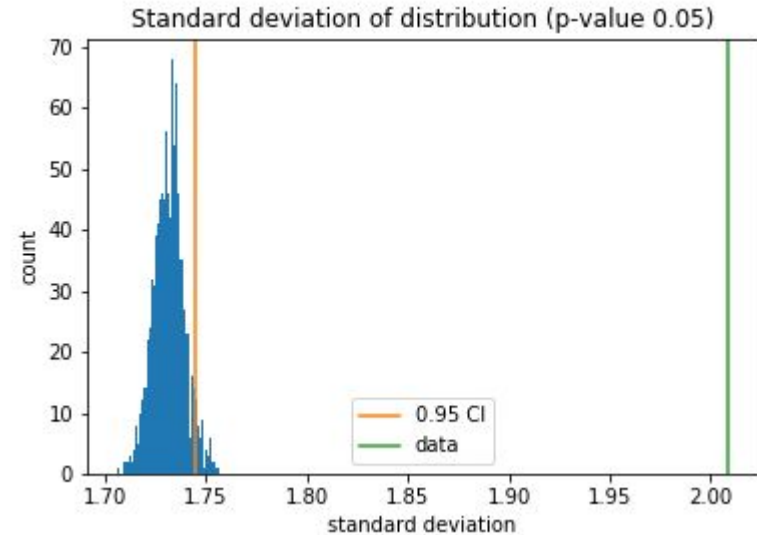
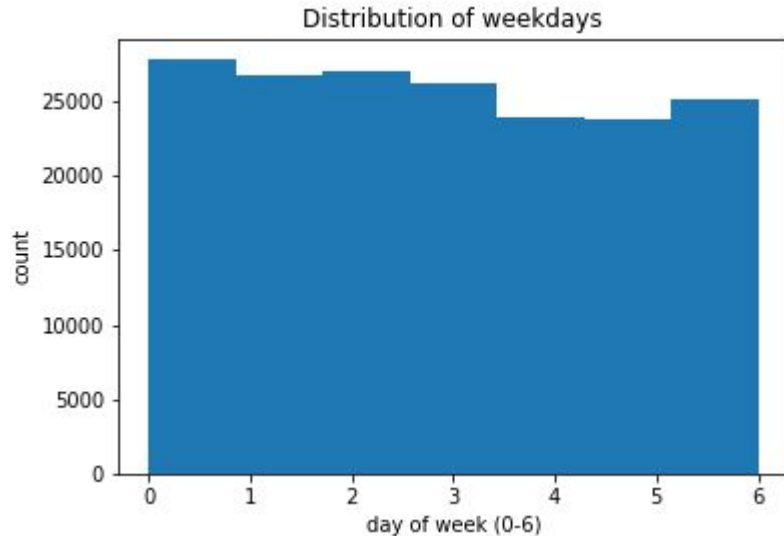
Uniformiteit Weekdagen

- Hypothese-test op σ van distributie, p-waarde 0.05



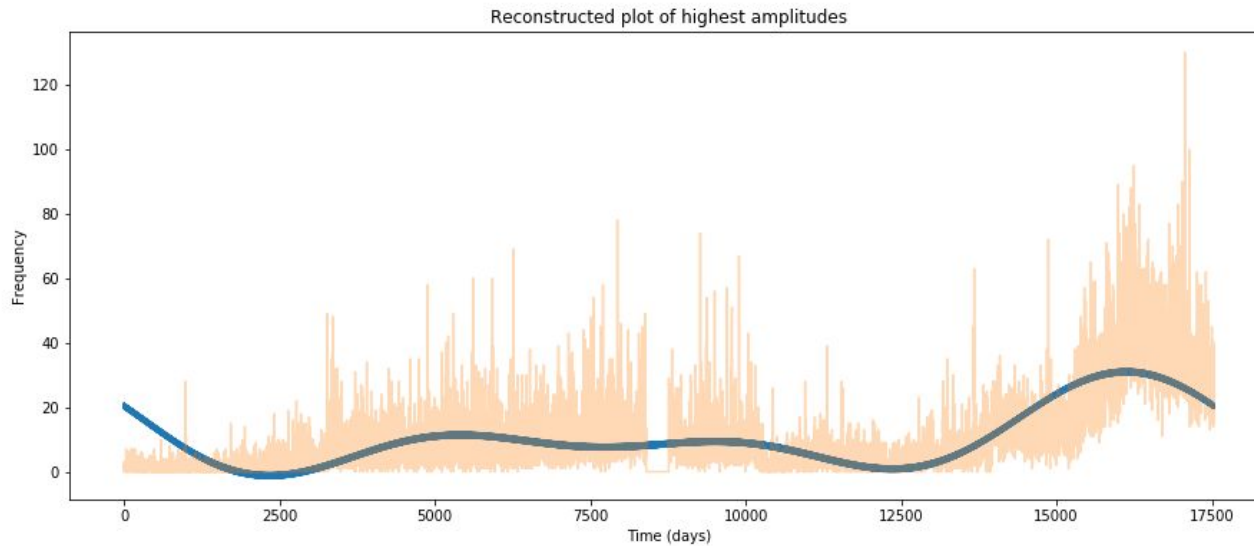
Uniformiteit Weekdagen

- Hypothese-test op σ van distributie, p-waarde 0.05



Fourier-Analyse - Resultaten

- IFFT top 9 amplitudes



Fourier-Analyse - Discussie

- Gat in data
 - FFT subset voor of na redelijk hetzelfde
- Niet in lijn met verwachtingen
- Primaire frequentie analyse niet waardevol

Time Predictor - Methode

- Hoe maak je een redelijke predictor voor de locaties en tijden van aanslagen?

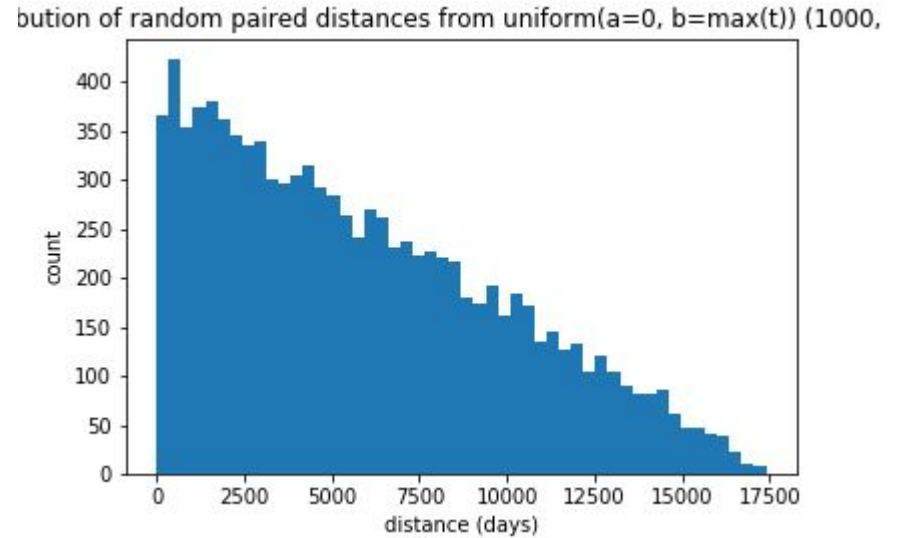
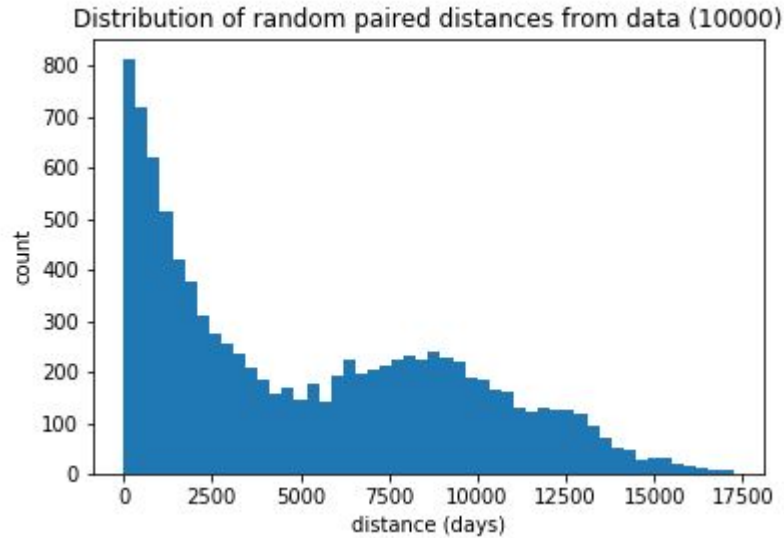
Time Predictor - Methode

- Hoe maak je een redelijke predictor voor de locaties en tijden van aanslagen?
 - Locatie aannemen en vervolgens aanslagen over tijd voorspellen
- Geïnteresseerd in plaatsen met veel aanslagen
- Plaatsnaam mist vaak in GTD-dataset

Time Predictor - Methode

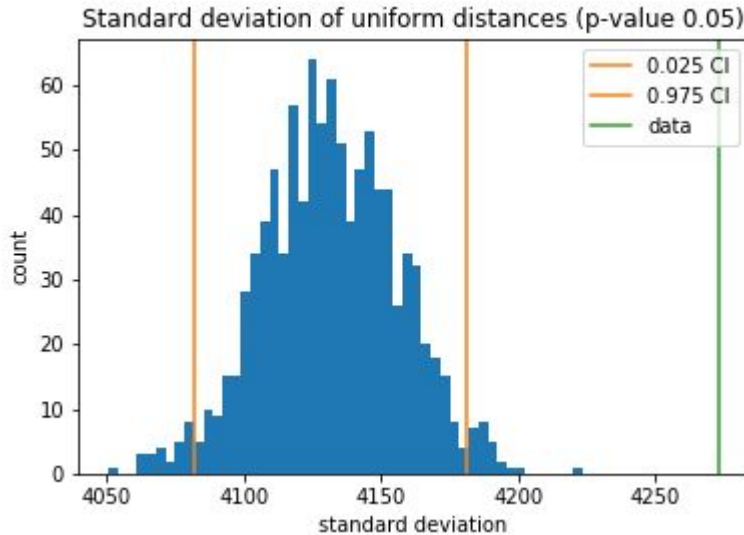
- Hoe maak je een redelijke predictor voor de locaties en tijden van aanslagen?
 - Locatie aannemen en vervolgens aanslagen over tijd voorspellen
- Geïnteresseerd in plaatsen met veel aanslagen
- Plaatsnaam mist vaak in GTD-dataset
 - Density-based clustering, cluster als locatie
- Time-series prediction van aanslagen over tijd in cluster

Tijdsafhankelijkheid



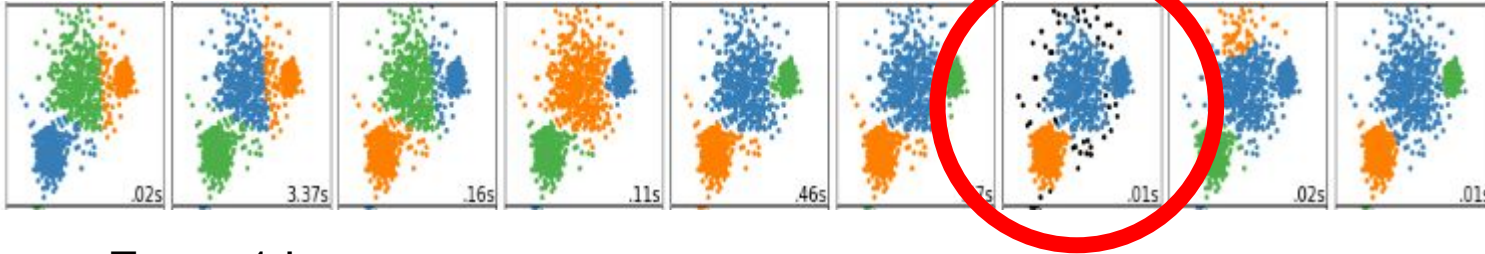
Tijdsafhankelijkheid

- Hypothese-test op σ van distributie van afstanden, p-waarde 0.05



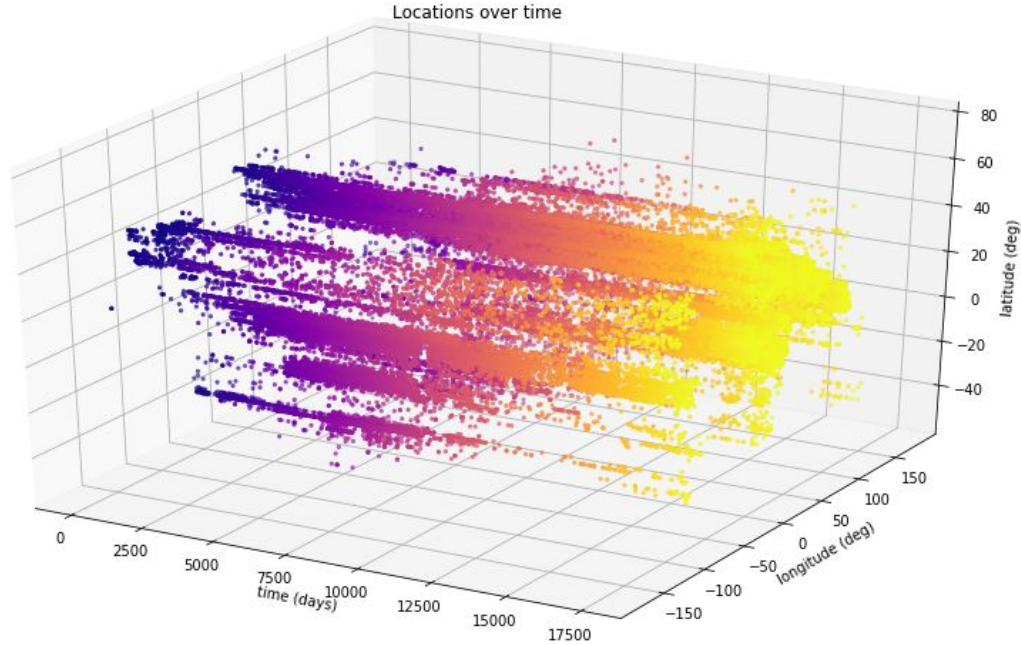
Time Predictor - Methode

- DBSCAN clustering
 - Density-based clustering
 - Allows unclustered 'noise'



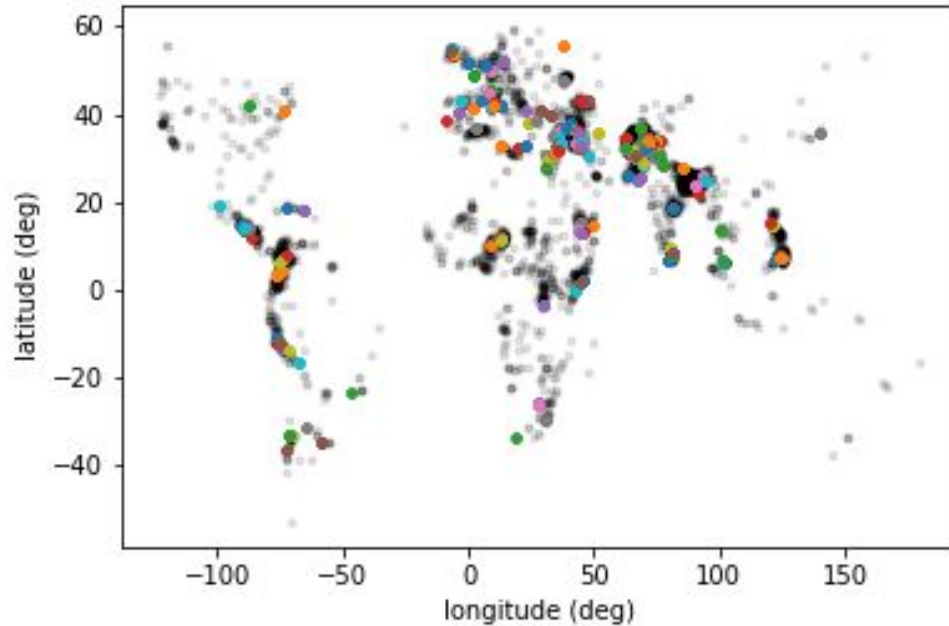
- Eps = 1 km
- Strided in implementatie, cluster = punten rond cluster mean
 - Rond cluster mean met $d = 10$ km

Time Predictor - Methode



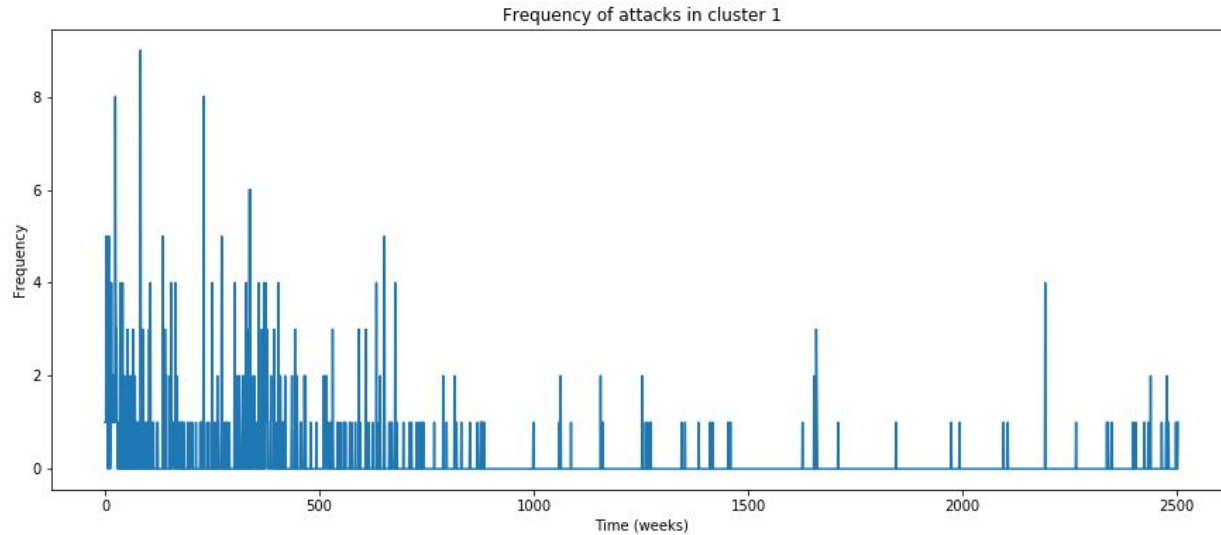
Time Predictor - Resultaten

ring of subset of locations (stride=20) (colored = clustered, gray = uncl



Time Predictor - Resultaten

- Voorbeeld frequentie in cluster



Time Predictor - Discussie

- Geen predictie...
 - Kan gedaan worden met LSTM RNN
 - Of (S-)ARIMA, misschien met 'seasonality' van de FFT

Time Predictor - Discussie

- Geen predictie...
 - Kan gedaan worden met LSTM RNN
 - Of (S-)ARIMA, misschien met 'seasonality' van de FFT
- Clustering komt ongeveer overeen met heatmap

Time Predictor - Discussie

- Geen predictie...
 - Kan gedaan worden met LSTM RNN
 - Of (S-)ARIMA, misschien met 'seasonality' van de FFT
- Clustering komt ongeveer overeen met heatmap
- Optimalisatie van DBSCAN parameters
 - Eps arbitrair gekozen door middel van 'testen'
 - Andere parameters ongezet

Time Predictor - Discussie

- Geen predictie...
 - Kan gedaan worden met LSTM RNN
 - Of (S-)ARIMA, misschien met 'seasonality' van de FFT
- Clustering komt ongeveer overeen met heatmap
- Optimalisatie van DBSCAN parameters
 - Eps arbitrair gekozen door middel van 'testen'
 - Andere parameters ongezet
- Clustering op subset
 - Redelijk stabiel en accuraat aan de hand van testen.
 - Nauwkeurigheid metriek implementeren

Conclusie en discussie

- Geen goede voorspeller, alleen weekdays significant
- GTD miste het jaartal 1993
 - Niet geheel onbelangrijk, o.a. de eerste aanslag op een van de WTC-gebouwen in NYC
 - Geen interpolatie
- Onafhankelijkheid aangenomen

Einde

Zijn er nog vragen?

Bronvermelding

1: <https://staff.fnwi.uva.nl/j.m.lagerberg/signaalverwerking/ctfs.pdf>