

Verslag Datavisualisatie: Groep 5

18-05-2024

Noah Vandamme

Simon Van den Bussche

Palle Measen

Onderzoeksvraag

We willen te weten komen of er een impact is op natuurrampen door de opwarming van de aarde (klimaatverandering). Dit kunnen we doen door de volgende vragen te beantwoorden:

- Is er een correlatie tussen de globale temperatuur en het aantal natuurrampen?
- Is er een correlatie tussen de globale temperatuur en het aantal doden?
- Is er een correlatie tussen de globale temperatuur en de krachtigheid van de natuurrampen? Merken we bijvoorbeeld aan aardbevingen dat deze een hogere waarde hebben op de schaal van richter? Staat meer gebied onder water bij overstromingen?

Dataset

We gebruiken de EM-DAT dataset[1]. Dit is de dataset van een organisatie die zich sinds 1988 bezig houdt met het verzamelen en rapporteren van alle rampen. Dit zijn zowel natuurlijke als onnatuurlijke (technologische) rampen. Over deze data worden er verschillende eigenschappen bijgehouden: het type van de ramp, welke regio(/land), de start & einddatum, het aantal doden/gewonden/getroffen personen, de schade (prijs), de magnitude, etc.. Over alle rampen is niet alle info beschikbaar, het is bijvoorbeeld mogelijk dat het aantal doden niet gekend is of de specifieke begin - en einddatum van de ramp.

Aangezien onze onderzoeksvraag gaat over natuurrampen filteren we de onnatuurlijke rampen uit deze dataset weg. De natuurrampen die we gaan bestuderen bestaan uit:

- overstromingen
- stormen
- aardbevingen
- droogtes
- bosbranden

EM-DAT stelt ook dat het dataset onderhevig is aan tijdsbias. Dit betekent dat het dataset lijdt aan ongelijke rapportagekwaliteit en dekkinggraad in de loop van de tijd. We kunnen deze bias verminderen door het dataset te filteren om alleen gegevens na het jaar 1988 op te nemen.

Naast de EM-DAT dataset gebruiken we ook de *NASA GISS Surface Temperature Analysis* [2] dataset voor de temperatuurdatta.

Indeling

We hebben ervoor gekozen om te werken observablehq. Dit gaf ons de tools om op een snelle manier de nodige grafieken te maken. We hebben een main page gemaakt om de trends over alle natuurrampen te bekijken. Vervolgens hebben we voor elke gekozen natuurramp een apparte page aangemaakt om meer specifieke trends te bekijken. We hebben enkele grafieken gemaakt die consistent waren voor alle pages. Vervolgens hebben we een paar grafieken die alleen op de main page staan, grafieken die enkel op de subpages staan. Tenslotte hebben we ook nog grafieken die op een enkele subpage staan.

Grafieken

Sunburst

We wouden een zicht krijgen hoe het aantal natuurrampen zich tot elkaar verhouden. Hiervoor creëerden we een bar chart in Figure 1 die het aantal natuurrampen per type weergaf. Dit was de eerste chart die we gemaakt hebben. Dit gaf een idee over de types rampen maar niet over hoe de subtypes van de rampen zich met elkaar verhouden.

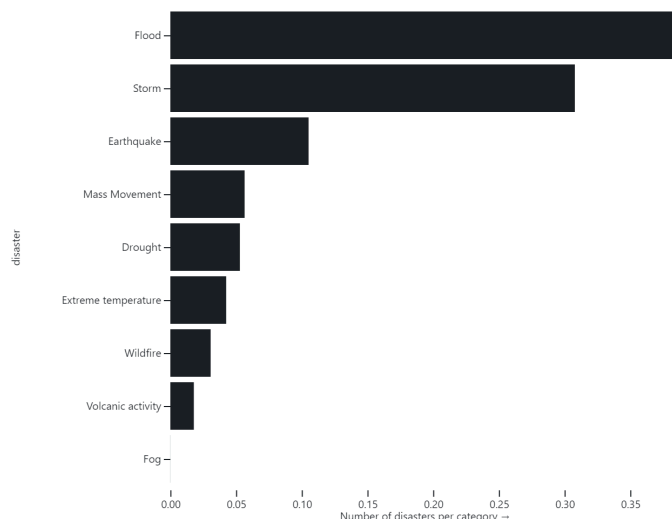


Figure 1: Bar chart met het aantal rampen per type

We hadden dus een grafiek nodig die de types en subtypes duidelijk toonde. Uiteindelijk hebben we gekozen voor de sunburst grafiek. De eerst implementatie is hieronder te zien. Op deze manier konden we de verschillende natuurrampen weergeven en het aandeel van elke natuurramp in de dataset. Om de defenitie van elk type weer te geven hebben we de definities van EM-DAT genomen en een feature toegevoegd waar de definities getoond worden als je over het type in de graph hovered.



Figure 2: Sunburst, eerste implemtatie

We kregen hier wel het probleem dat een aantal subtypes een veel te klein aandeel hadden om deze duidelijk te weergeven in de sunburst.



Figure 3: Sunburst, Subtypes met een te klein aandeel

Dit hebben we uiteindelijk opgelost door een categorie: other te maken en de subtypes met een te klein aandeel in deze categorie te zetten. Het uiteindelijke reesultaat is te zien in Figure 4.



Figure 4: Sunburst, Finale implementatie

We hadden ook gekeken naar de TreeMap grafiek om onze dataset weer te geven maar vonden dat deze onduidelijker was dan de sunburst.

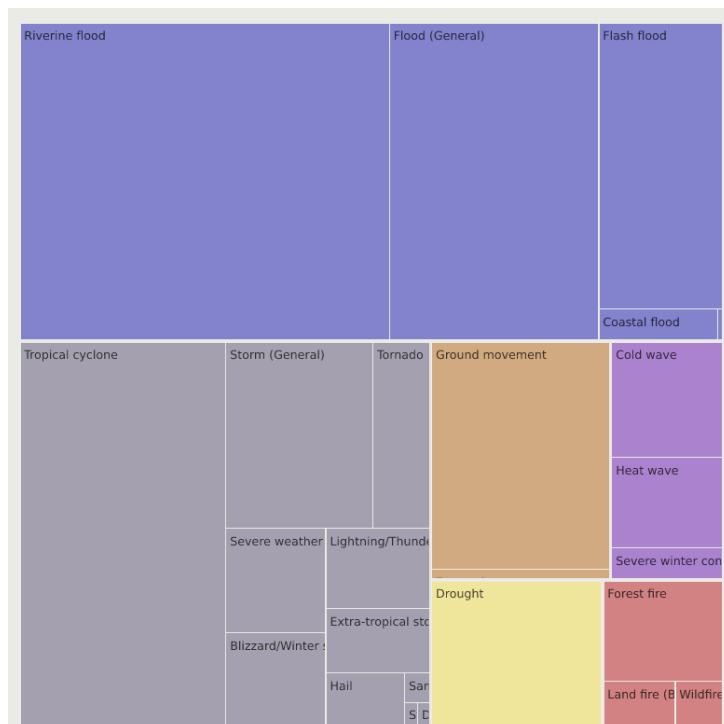


Figure 5: TreeMap

Trend area chart

Aangezien we de trend wouden bekijken van de natuurrampen kregen we direct het idee 1 grafiek centraal te zetten die de trend in aantal natuurrampen weergaf. Het eerste idee was om een bumpchart te maken om de ranking van de natuurrampen weer te geven. Het resultaat is te zien in Figure 6.

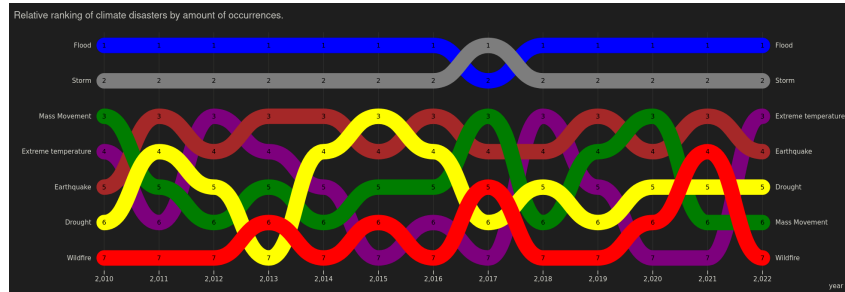


Figure 6: Bumpchart met de ranking tussen de verschillende natuurrampen

Na het maken van deze figuur kregen we het idee dat deze niet veel bijdraagt. Het geeft de relatieve volgorde mee maar niet de globale trend. Daarom hebben we een line chart gemaakt die te zien is in Figure 7. Hoewel de relatieve ranking duidelijk is is het nog steeds moeilijk om een globale trend te zien.

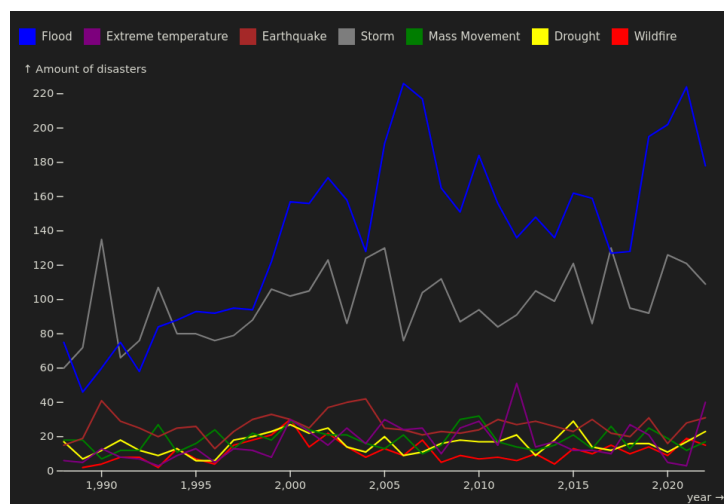


Figure 7: Area chart met global trend van natuurrampen

Dit hebben we verholpen door het in een area chart te veranderen. Deze is te zien in Figure 8. De area chart geeft beide de ranking aan tussen de natuurrampen en de globale trend.

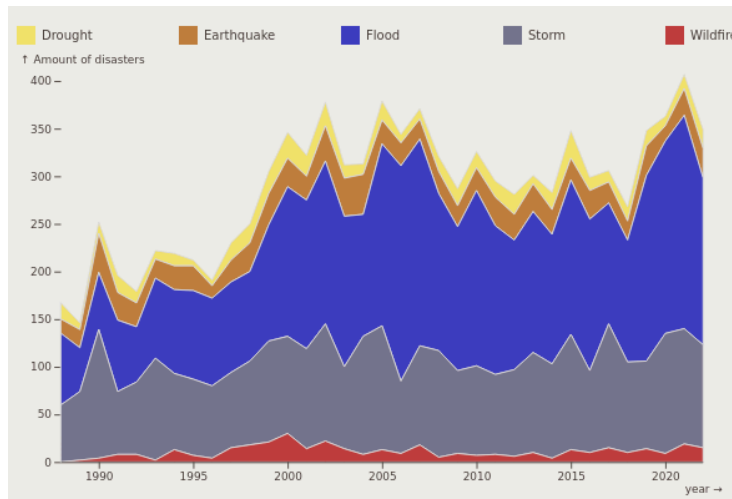


Figure 8: Area line met trend per natuurramp

Death toll bar chart

Om weer te geven welke soorten natuurrampen het meeste impact hadden op de wereld hebben we een bar chart gemaakt die het totaal aantal doden per natuurramp meeg geeft. Deze is te zien in Figure 9.

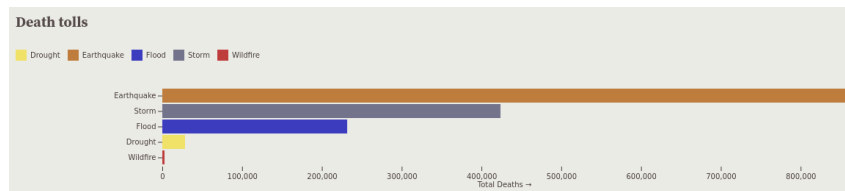


Figure 9: Bar chart met dodentallen per natuurramp

Dodelijkste natuurrampen

We wouden ook kijken of er een trend zat in hoeveel mensen er werden beïnvloed door de natuurrampen. Dit gaf ons de grafiek in Figure 10. Hier is bijna geen trend te zien. Dit komt vooral omdat in de dataset, de kolom met *people affected* veel lege entries heeft. Idem met Deaths en Injuries. Hoewel we de graaf niet konden gebruiken gaf het ons wel een idee. Aangezien er vaak uitschieters waren, waren we vaak nieuwsgierig naar welke natuurrampen deze uitschieters veroorzaakten. Daarom hebben we de grafiek in Figure 11 gemaakt. Deze bar chart geeft de meest dodelijke natuurrampen weer. Ook hebben we een manier gemaakt om te filteren per land.

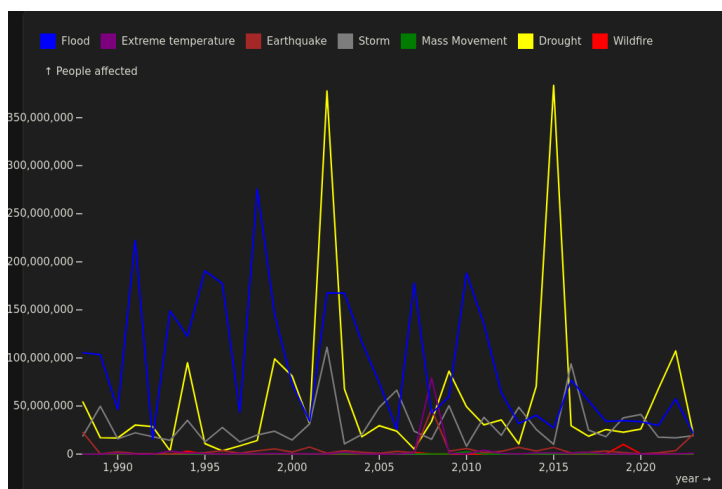


Figure 10: Line chart met People affected

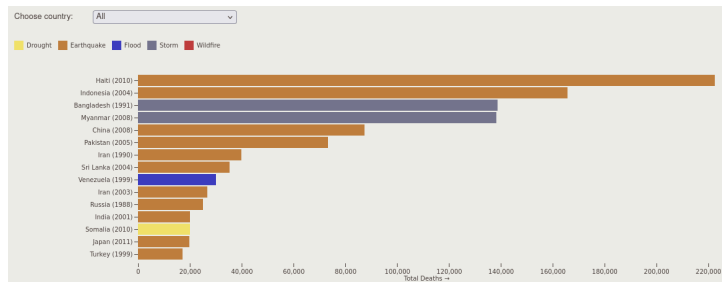


Figure 11: Bar chart met dodelijkste natuurrampen

Correlatie met temperature

Om de correlatie tussen de opwarming van de aarde en het aantal natuurrampen te bekijken hadden we eerst het idee om een line chart te nemen van de temperatuur en deze naast de trend are chart te zetten. het resultaat was de grafiek hieronder.

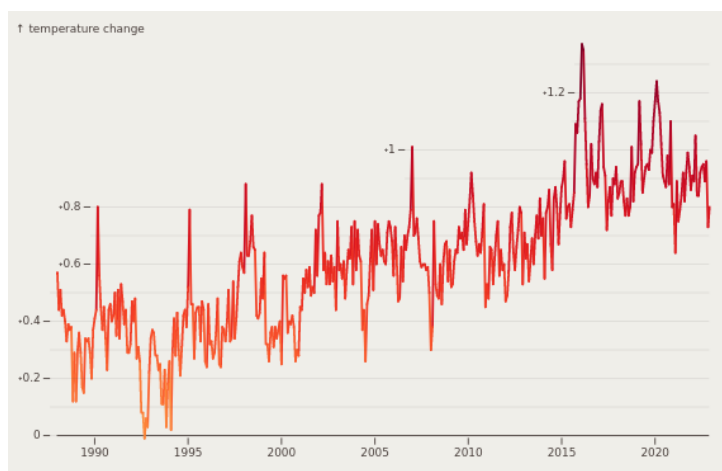


Figure 12: Line chart met global temperature anomalies

We hadden door dat het vrij moeilijk was om de trends objectief met elkaar te kunnen vergelijken. Het idee was een nieuwe graph te maken met beide de trend van het aantal natuurrampen en de temperatuur anomaly. Het resultaat is hieronder te zien. Om de correlatie minder subjectief te maken hebben we ook een correlatie-factor in de graph geplaatst.

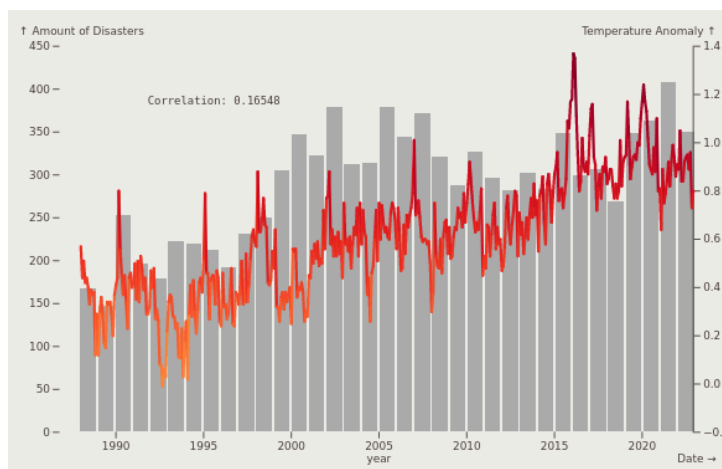


Figure 13: Line chart met global temperature anomalies en bar chart met aantal natuurrampen per jaar

Wereldkaarten

Om een idee te geven waar de meeste natuurrampen van een specifiek type plaatsvinden hebben we een wereldkaart gemaakt. Deze is te zien in Figure 14. Aangezien er enkele landen zijn die veel meer entries hebben van een bepaald natuurramp-type dan het gemiddelde, is het vrij moeilijk om het verschil te zien tussen landen met een normaal aantal natuurrampen. Dit is waarom we de optie hebben toegevoegd om de kaart met een logaritmische schaal te bekijken.

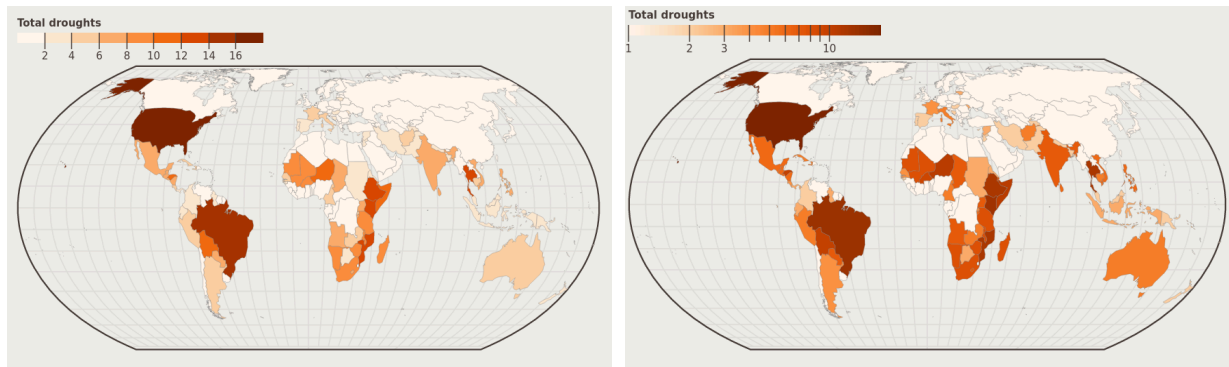


Figure 14: Kaart met aantal natuurrampen per land: *Links*: lineair, *Rechts*: logaritmisch

Tenslotte om er voor te zorgen dat we ook een verschil zien tussen kleinere landen hebben we een full-view wereldkaart toegevoegd. Deze is te zien in Figure 15

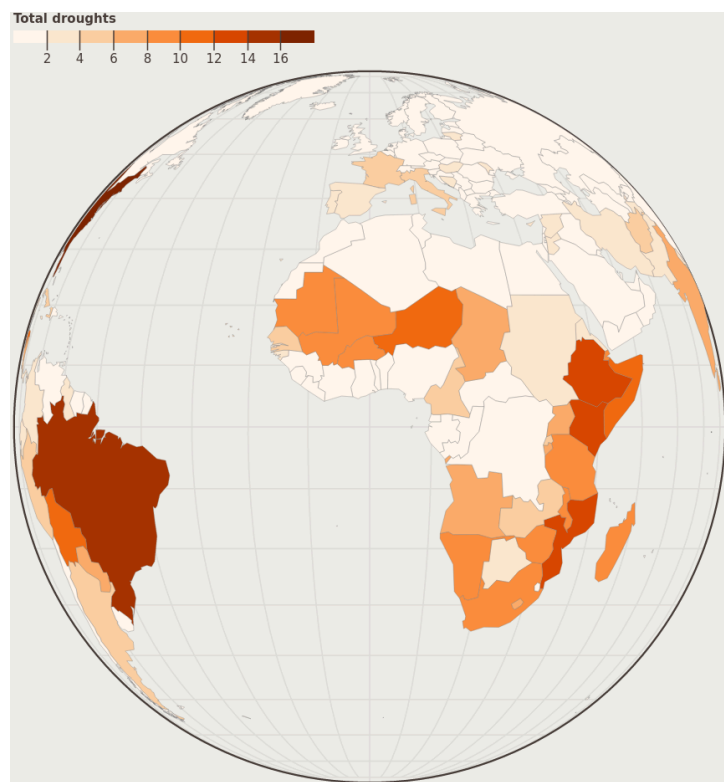


Figure 15: Ingezoomde kaart met aantal natuurrampen per land

Natuurramp specifieke grafieken

Naast de bovenstaande grafieken hebben we ook nog enkele grafieken specifiek voor een natuurramp-type.

Scatter plot droogte en bisbranden duur

Om te kijken of er een correlatie is tussen de opwarming van de aarde en de duur van droogtes en bosbranden, hebben we de scatter charts in Figure 16 gemaakt. Voor de droogtes geeft deze een duidelijk stijgende lijn weer. Idem voor de duur van bosbranden.

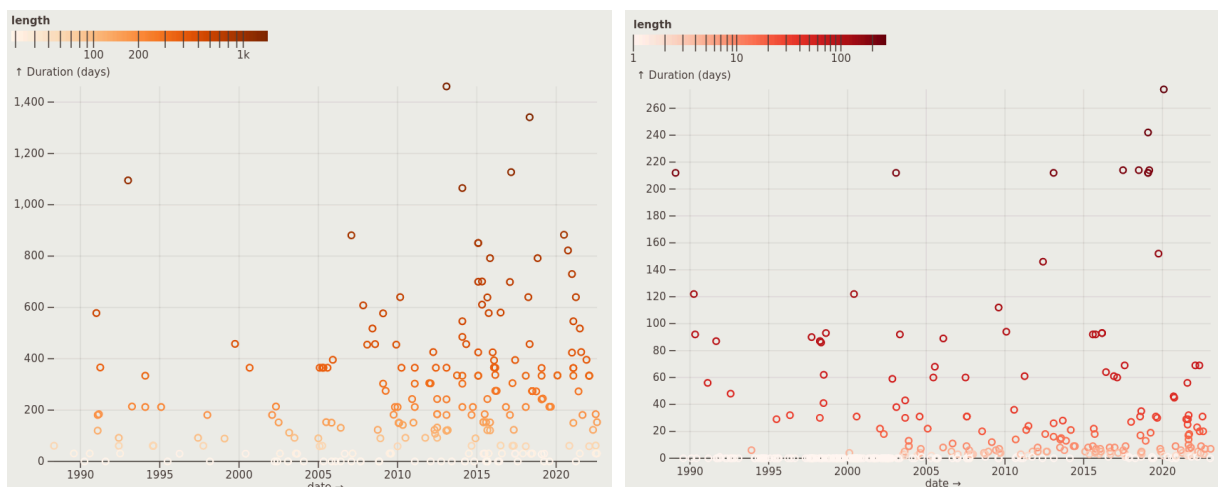


Figure 16: Scatter plot met de duur van natuurrampen: *Links*: Droogtes, *Rechts*: bosbranden

Scatter plot grootte overstroomd gebied

Hetzelfde idee als voor de droogtes en bosbranden hadden we voor de grootte van de overstroomde gebieden. Hier kregen we wel het probleem dat er geen duidelijke stijgende lijn was. Daarom hebben we linear regression lijn getrokken door de data om een duidelijk stijgende trend weer te geven.

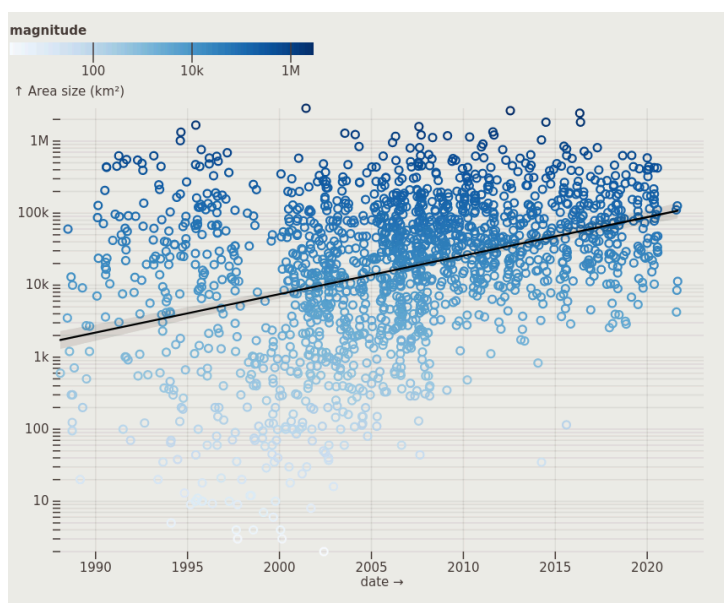


Figure 17: Een scatter plot met de grootte van de overstroomde gebieden

Het was ook de bedoeling dit te doen voor de oppervlakte van branden maar er was te weinig data over de grootte van het gebied in de dataset.

Aarbevingen kaart

Voor de aardbevingen leek het ons ook interessant om te kijken waar de zwaarste aardbevingen plaatsvinden. Aangezien de aardbevingen exacte coördinaten bevatten in de dataset konden we vrij specifiek aanduiden waar de zwaarste aardbevingen zich afspeelden.

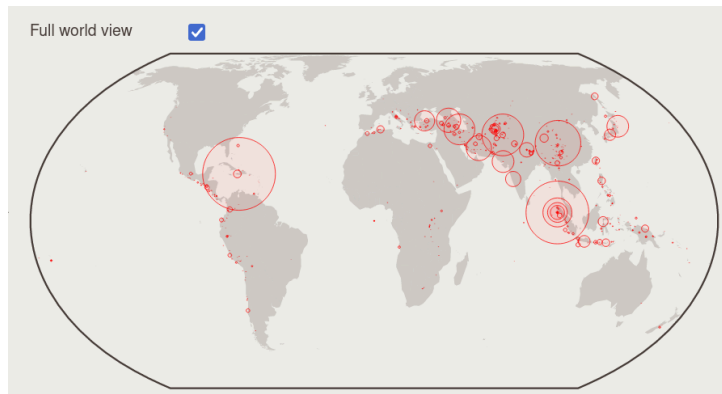


Figure 18: Een kaart met de intensiteit van aardbevingen

Ideeen niet in het resultaat

Correlatie matrix

We hadden het idee om te kijken of er een correlatie was tussen de verschillende subtypes van natuurrampen. Hiervoor hadden we een correlatieMatrix gemaakt. Deze is te zien in Figure 19. Aangezien er vrij weinig informatie uit de matrix te halen is hebben we deze weggelaten.

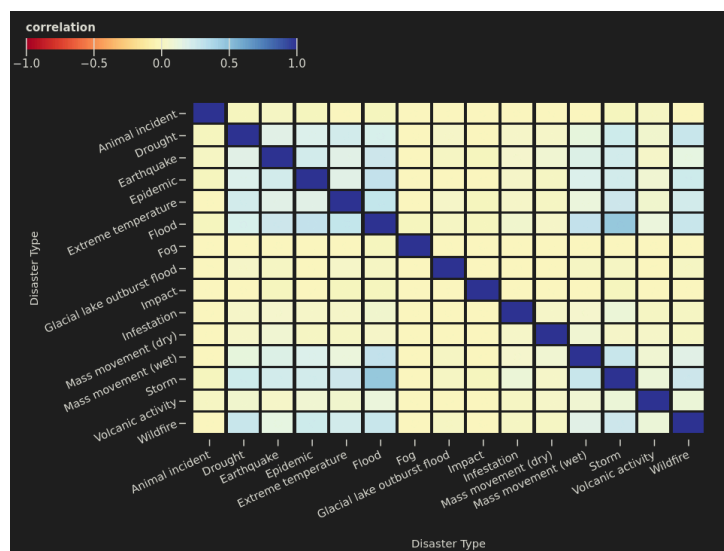


Figure 19: Correlatie matrix

Bibliografie

- [1] EM-DAT, "EM-DAT: The International Disaster Database." 2008.
- [2] "GISS surface temperature analysis (GISTEMP v4)."