

# Ruimtelijke datavisualisatie

## Een verkenning van verschillende datavisualisatie methoden in GIS

**Menno ter Bals**  
**Leander Kuijvenhoven**

### **Inleiding**

*Voor organisaties als statistische bureaus die zich bezig houden met het verzamelen, verwerken en analyseren van data en het publiceren van de daaruit voortkomende resultaten bestaat een breed en beproefd scala aan onderzoeksmethoden en publicatiemogelijkheden. Een belangrijke component binnen dit geheel is de ruimtelijke dimensie. Wanneer data een ruimtelijke component bevat, meestal in de vorm van een geografische locatie, wordt vaak van 'geo-informatie' gesproken. Geo-informatie is geschikt voor ruimtelijke analyse. Ruimtelijke analyse biedt onder andere de mogelijkheid data op basis van locatie te koppelen aan andere gegevens, de data visueel in kaart te brengen en op basis hiervan ruimtelijke clusters en ruimtelijke verbanden te herkennen en in kaart te brengen. Een goed voorbeeld van geo-informatie is de door het CBS verzamelde censusdata. Voor ieder huishouden in de laatste census van 2011 is een geografische locatie, in de vorm van coördinaten, geregistreerd. Dit betekent dat deze locatie informatie te koppelen is aan alle personen binnen deze huishoudens en dat alle eigenschappen van huishoudens en personen, welke geregistreerd zijn in de census, ruimtelijk geanalyseerd en in kaart gebracht kunnen worden.*

*In dit artikel wordt een verkenning gemaakt van enkele verschillende mogelijkheden op het gebied van ruimtelijke datavisualisatie. Met behulp van geo-informatie uit de Curaçaose census van 2011 worden enkele visualisatie methoden geïllustreerd en met elkaar vergeleken.*

### **Data en methodologie**

In dit artikel wordt gebruik gemaakt van census data van 2011. Een groot voordeel bij het gebruik van censusdata is dat gedetailleerde analyses, bijvoorbeeld statistieken op buurtniveau, mogelijk zijn gezien het omvattende karakter van censusdata. Een ander voordeel is dat speciale (soms zeldzame) gebeurtenissen gedetecteerd worden, die bij steekproeven vaak niet waargenomen worden. Bovendien is het een uitstekende manier om de kracht van visualaties aan te tonen.

Volgens Schmid (1983) zijn er de volgende voordelen aan een (goed gemaakte) visualisatie:

1. De kracht van visualisaties is dat de kerngedachte van de gegevens snel en vrijwel direct kan worden overgebracht.
2. Een visualisatie is een krachtig instrument om een veelomvattend en samenhangend beeld te geven van de gegevens
3. Een visualisatie heeft veelal meer overtuigingskracht dan een stuk tekst of een tabel.
4. Een visualisatie kan eerder onbekende patronen bloot leggen.
5. Een visualisatie trekt eerder de aandacht van de lezer door zijn aantrekkelijke uiterlijk vergeleken met tekst of tabellen.

In dit artikel zal een diversiteitsindex gebruikt worden als statistiek om de verschillende visualisatietechnieken te illustreren. Een diversiteitsindex is een maat die op basis van verschillende groepen aangeeft op welke wijze en in welke mate de groepen verdeeld zijn in een bepaald gebied. Hoewel er verschillende soorten diversiteitsindices zijn, wordt hier de Shannon-Weiner Index gebruikt. Reard en Firebaugh (2002) vergeleken zes verschillende diversiteitsindices die als doel hadden sociale verschijnselen te kwantificeren. De Shannon-Weiner Index kwam hier als beste diversiteitindex naar voren. In dit artikel wordt deze index gebruikt. In wat volgt zal met de diversiteitsindex de Shannon-Weiner Index bedoeld worden. De reden hiervoor is dat het bij de term diversiteitsindex direct duidelijk is dat het om het meten van diversiteit gaat. Bovendien kan er geen verwarring ontstaan over welke diversiteitsindex wordt bedoeld, omdat er hier slechts één index wordt gebruikt.

De diversiteitsindex wordt hier toegepast op twee groepen van personen. De eerste groep bestaat uit bewoners die geboren zijn in Curaçao en de tweede groep bestaat uit bewoners die elders zijn geboren. De diversiteitsindex meet dus de mate van diversiteit in een bepaald gebied, gemeten door naar geboorteland van de inwoners te kijken. Per gebied wordt de proportie mensen die geboren zijn in Curaçao berekend en de proportie van mensen die niet in Curaçao geboren zijn berekend. Deze proporties vormen de basis voor de berekening van de diversiteitsindex. Deze diversiteitsindex loopt van 0 (lage diversiteit) tot en met 1 (hoge diversiteit)<sup>17</sup>.

De visualisaties in dit artikel zijn niet bepaald geschikt voor het afdrukken op papier in grijstinten. Het gebruik van kleur is juist een belangrijk functioneel middel binnen het gebied van geografische visualisatie. Daarbij is de kaartschaal waarop gepubliceerd kan worden in een uitgave op A4 formaat gelimiteerd. Daarom is voor dit artikel gebruik gemaakt van de mogelijkheid met GIS om online digitale kaarten te publiceren. Dat wil zeggen dat de visualisaties (kaarten) uit dit artikel online te bezichtigen zijn via een link die duidelijk staat aangegeven op de CBS website ([www.cbs.cw](http://www.cbs.cw))<sup>18</sup>. De hier weergegeven kaarten in grijstinten zijn daarom slechts beperkte voorbeelden.

---

<sup>17</sup> Bij de hier gebruikte diversiteitsindex wordt eerst  $H = -\sum_{i=1}^2 p_i \ln p_i$ , waarbij  $p_i$  de proportie aangeeft van groep  $i$ , berekend. Daarna wordt de uiteindelijke diversiteitsindex berekend  $H/\ln 2$ . Op deze wijze loopt de index van 0 en 1.

<sup>18</sup> Link naar web applicatie met visualisaties:

<http://cbscuracao.maps.arcgis.com/apps/MapSeries/index.html?appid=d3382904c4074ce5994c95a2cedf845c>

## Geografische Informatiesystemen

Een geografisch informatiesysteem, afgekort GIS, is een geo-informatiesysteem waarmee ruimtelijke data kan worden verzameld, bewerkt, beheerd, geïntegreerd, geanalyseerd en gepresenteerd. In de praktijk verwijst een GIS vaak naar een geautomatiseerd systeem waarin deze genoemde functionaliteit gecombineerd is. Het CBS Curaçao maakt gebruik van ArcGIS. Dit is een GIS platform waarin verschillende software applicaties zijn gecombineerd om de volledige functionaliteit van een GIS te bieden (ESRI, 2016).

Voor de visualisaties en statistische analyse in dit artikel is gebruik gemaakt van de ArcGIS Desktop en ArcGIS Pro software. De online publicatie van de visualisaties is door middel van de cloud-based service ArcGIS Online tot stand gebracht.

## Visualisatie methoden

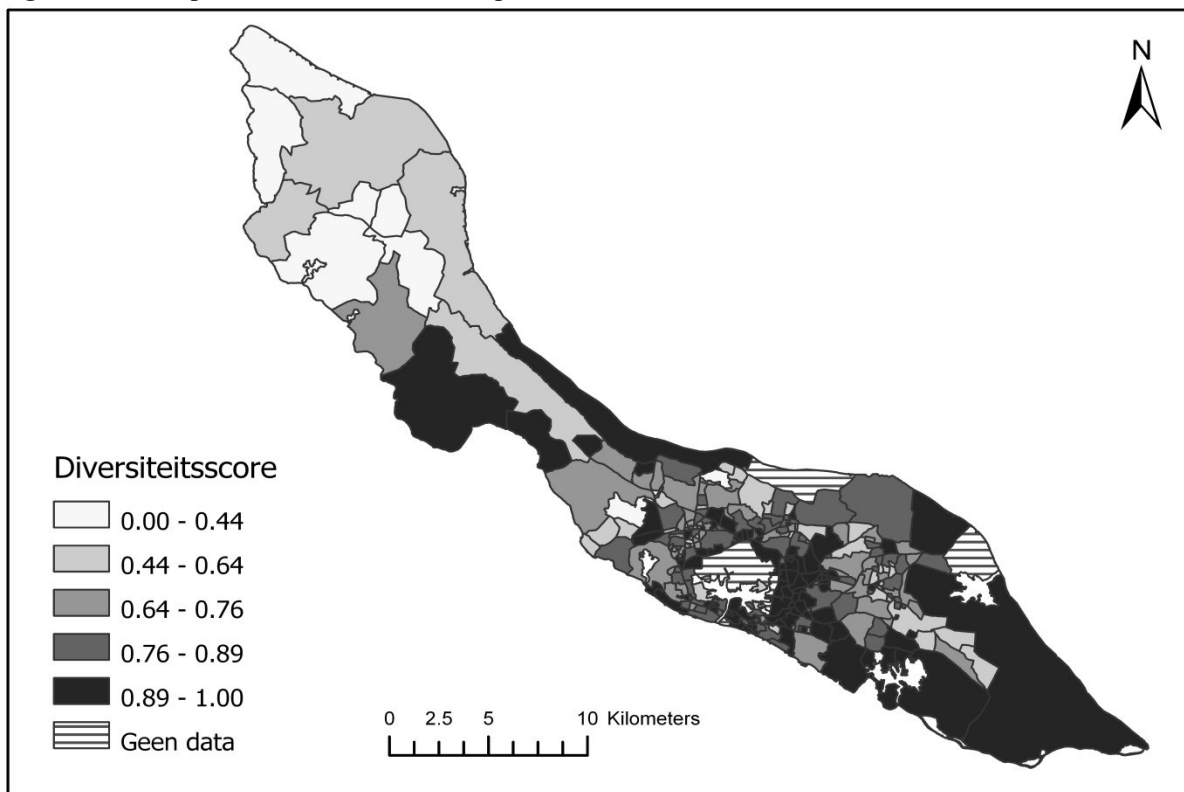
De data die verzameld en geanalyseerd wordt door het CBS leent zich bij uitstek tot het maken van thematische kaarten. Een thematische kaart is een visualisatie waarbij de verdeling van een variabele of combinatie van variabelen over een gebied wordt weergegeven. Bijvoorbeeld sociaal-economische verschijnselen, gekwantificeerd in indicatoren kunnen in kaart worden gebracht.

Er zijn verschillende soorten visualisatietechnieken om deze indicatoren te visualiseren. Achtereenvolgens zullen in deze paragraaf visualisaties door middel van een choropleet en door middel van rasters (vierkanten en hexagonalen) worden toegelicht.

## Choropleet

Een choropleet geeft de waarde van een bepaalde indicator per gebied aan door middel van kleurtinten. De gebieden zijn vaak administratieve gebieden. De kleurtint representeert de hoogte van de waarde van de indicator in een gebied. Er zijn verschillende classificatiemethoden om een gebied, op basis van een waarde, van een bepaalde kleurtint te voorzien. Voorbeelden zijn classificaties op basis van kwantielscores, gelijke intervallen, geometrische progressies en deviatie scores. In dit artikel wordt de veel gebruikte Jenks natural breaks classificatie methode gebruikt. Deze methode probeert de spreiding (uitgedrukt in variantie) van de waardes binnen een klasse of gebied klein te houden, terwijl de spreiding tussen gebieden zo groot mogelijk wordt gehouden. Op deze manier zijn de waardes binnen een gebied zo homogeen mogelijk, terwijl de waardes van de gebieden onderling zo verschillend mogelijk zijn, d.w.z. zo heterogeen mogelijk zijn

**Figuur 1. Choropleet – Diversiteitsindex per buurt, Curaçao, Census 2011**



In figuur 1 wordt de diversiteitsindex per buurt in Curaçao (Census 2011) weergegeven. De grijs tinten lopen van licht naar donker en de bijbehorende diversiteitsindex van laag (0) naar hoog (1). Met andere woorden, hoe donkerder de kleur in een buurt, hoe hoger de diversiteitsindex. Hoge diversiteitsindices duiden een meer heterogene samenstelling van de bevolking in een buurt aan, terwijl lagere diversiteitsindices op een meer homogene samenstelling van de bevolking duiden. Belangrijk is op te merken dat een hoge diversiteitsindex niet hoeft te duiden op sociale of culturele uitwisseling tussen de verschillende groepen.

In figuur 1 is te zien dat de bevolking in Ser'i Papaya een diversiteitsindex heeft tussen 0,00 en 0,44 (precieze waarde 0,36); een relatief lage mate van diversiteit op het gebied van geboorteland van de inwoners. Groot Davelaar laat daarentegen een diversiteitsindex tussen de 0,89 en 1,00 zien (precieze waarde 0,97); een relatief grote mate van gemengdheid op het gebied van geboorteland van de inwoners. Voor de totale bevolking van Curaçao is de diversiteitsindex 0,55. Deze score is gebaseerd op het gegeven dat 76 procent van de bevolking in Curaçao geboren is tegen 24 procent van de bevolking die elders geboren is.

Een choropleet met een groot aantal polygonen (buurten in dit geval), welke in oppervlakte sterk variëren, is niet erg geschikt om een gedetailleerde weergave te bieden van een thema wanneer deze kaart op klein formaat wordt afgedrukt. De kleine buurten zijn dan nauwelijks te onderscheiden in de

kaart. Daarbij kan er in een choropleet vertekening optreden wanneer niet alle gebieden dezelfde grootte of oppervlakte hebben. Dit wordt in het Engels aangeduid met *area bias*. Om het effect van area bias te elimineren kan er gebruik gemaakt worden van normalisering<sup>19</sup> of van rasters, waarover in de volgende paragraaf meer. De choropleet impliceert tevens een homogene verdeling binnen de individuele vlakken van het in kaart gebrachte fenomeen. Deze veronderstelling is, met name in grotere gebieden, vaak onjuist. In figuur 1 is het tevens lastig om de verschillende grijsinten te kunnen onderscheiden. Wel geeft de kaart een globaal overzicht van de ruimtelijke spreiding van de indicator over het gehele studiegebied.

Opgemerkt moet worden dat er gebieden zijn waar er geen data beschikbaar was om de indicator te berekenen. In dit voorbeeld had dit zelden te maken met non-respons, maar waren deze gebieden vaak onbewoond. Deze gebieden zijn gearceerd weergegeven in figuur 1.

## Vierkante en hexagonale rasters

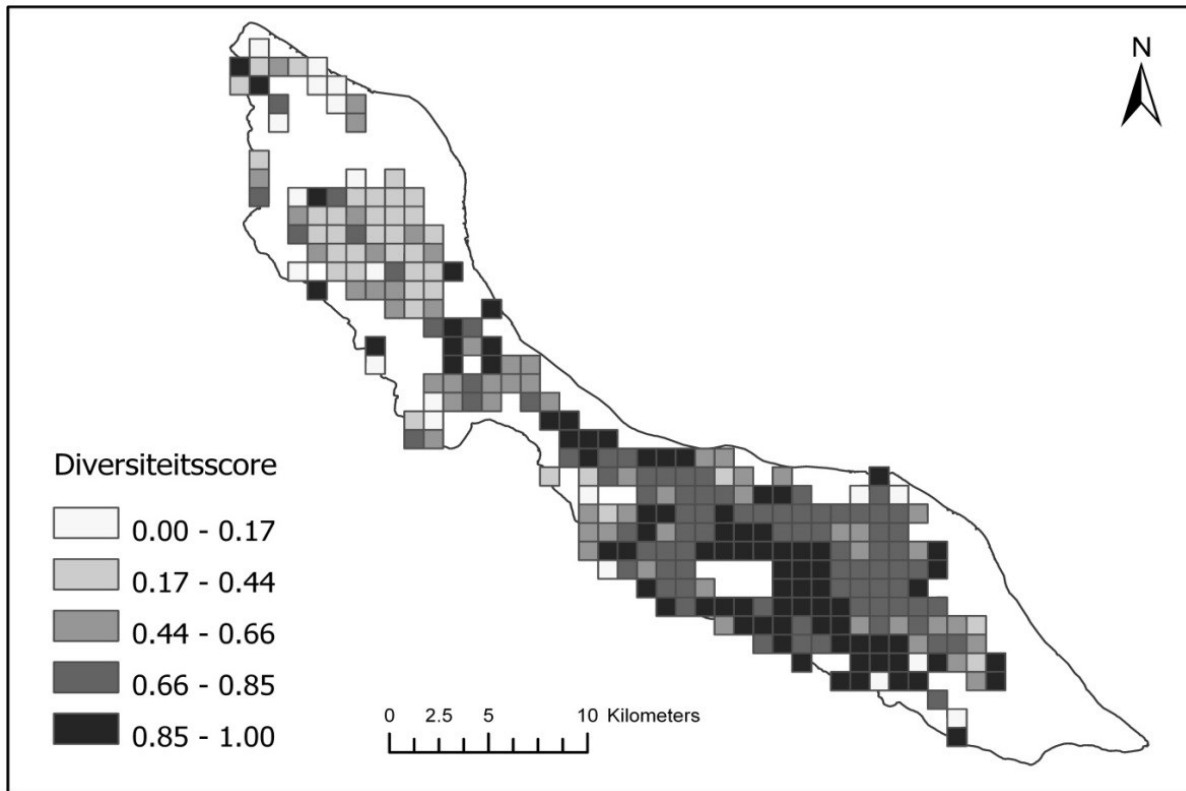
In dit artikel worden twee type rasters belicht: vierkante en hexagonale rasters. Bij het gebruik van rasters wordt het totale studiegebied onderverdeeld in gebieden van gelijke vorm en oppervlakte. De vraag die opkomt is op welke wijze het totale oppervlakte of gebied opgedeeld dient te worden. Buiten het vaak gebruikte vierkante raster zijn er ook andere mogelijkheden. Een van de criteria zou kunnen zijn dat het raster het oppervlakte zo economisch mogelijk indeelt. In dit geval moeten de gebieden op basis van gelijke oppervlaktes en een zo klein mogelijke omtrek worden ingedeeld. Het blijkt dat een honinggraat of hexagonaal raster als enige raster aan dit criteria voordoet. Figuur 2 en 3 zijn voorbeelden waarbij gebruik is gemaakt van een raster van vierkanten en figuur 4 en 5 zijn voorbeelden van hexagonale rasters. Voor beide raster soorten is op basis van alle datapunten binnen elke van de vierkanten of hexagonalen de diversiteitsindex per vierkant of hexagonaal berekend<sup>20</sup>. Bovendien zijn de vierkanten en hexagonalen waarover geen data beschikbaar was in het geheel weggelaten, terwijl de contour van Curaçao behouden is.

---

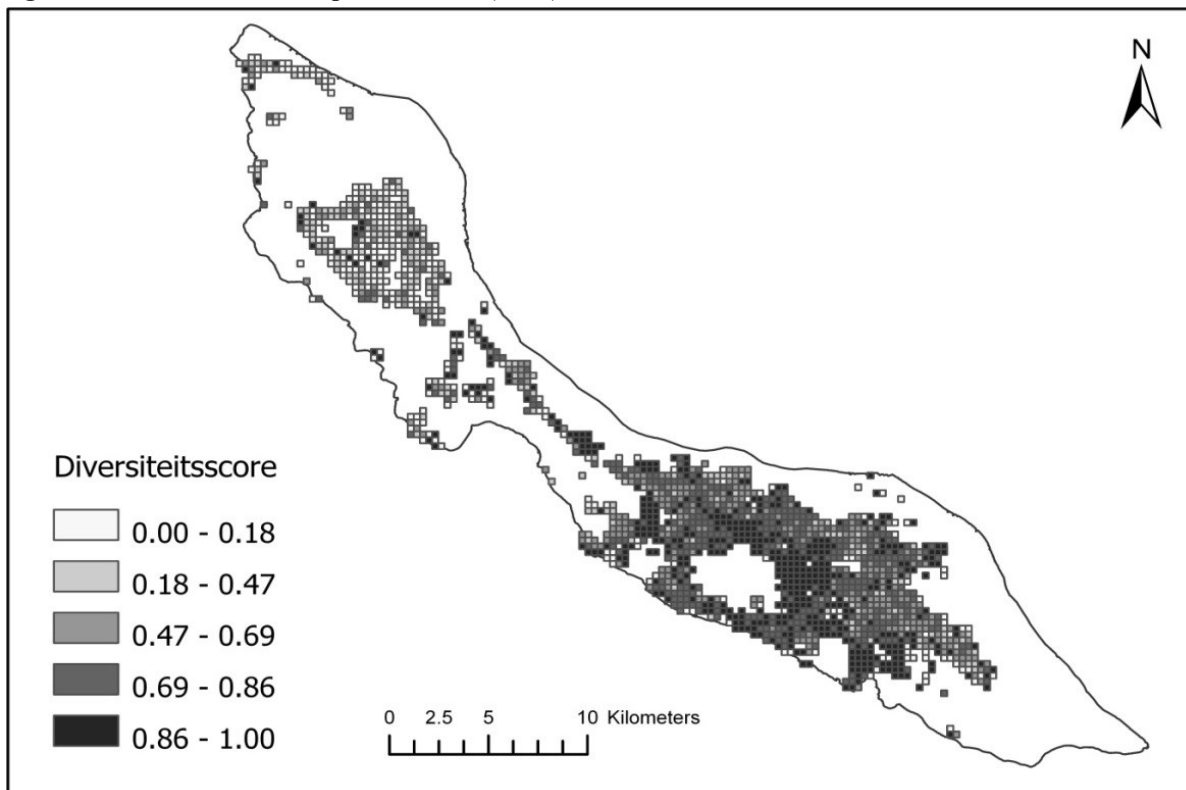
<sup>19</sup> Wanneer het om een *aantal* waarnemingen per gebied gaat dient te worden genormaliseerd naar bijvoorbeeld oppervlakte van het gebied. Het aantal waarnemingen per gebied wordt dan gedeeld door de oppervlakte van het gebied om tot de genormaliseerde waarde te komen. Op deze manier wordt een relatieve waarde verkregen die vergeleken kan worden met de relatieve waarden in de andere gebieden in de kaart.

<sup>20</sup> Zie ook de online visualisaties waarin door middel van zogenaamde ‘swipe’ kaarten de verschillende rasters direct met elkaar vergeleken kunnen worden.

**Figuur 2. Diversiteitsindex per vierkant (100ha), Curaçao, Censüs 2011**



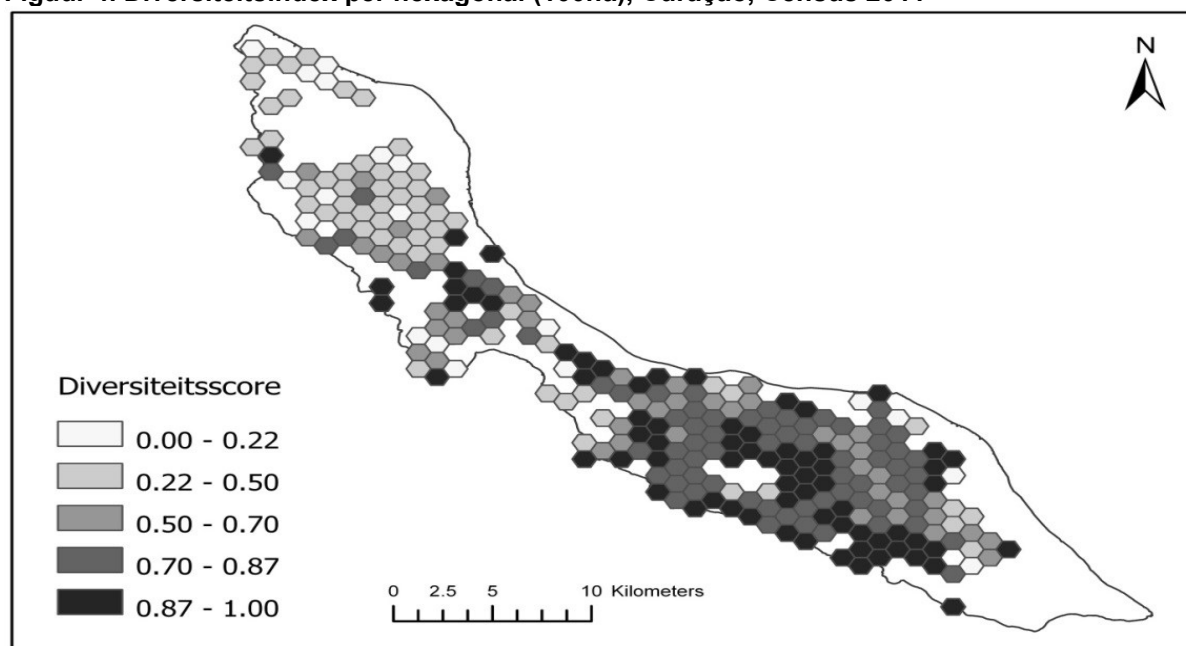
**Figuur 3. Diversiteitsindex per vierkant (10ha), Curaçao, Censüs 2011**



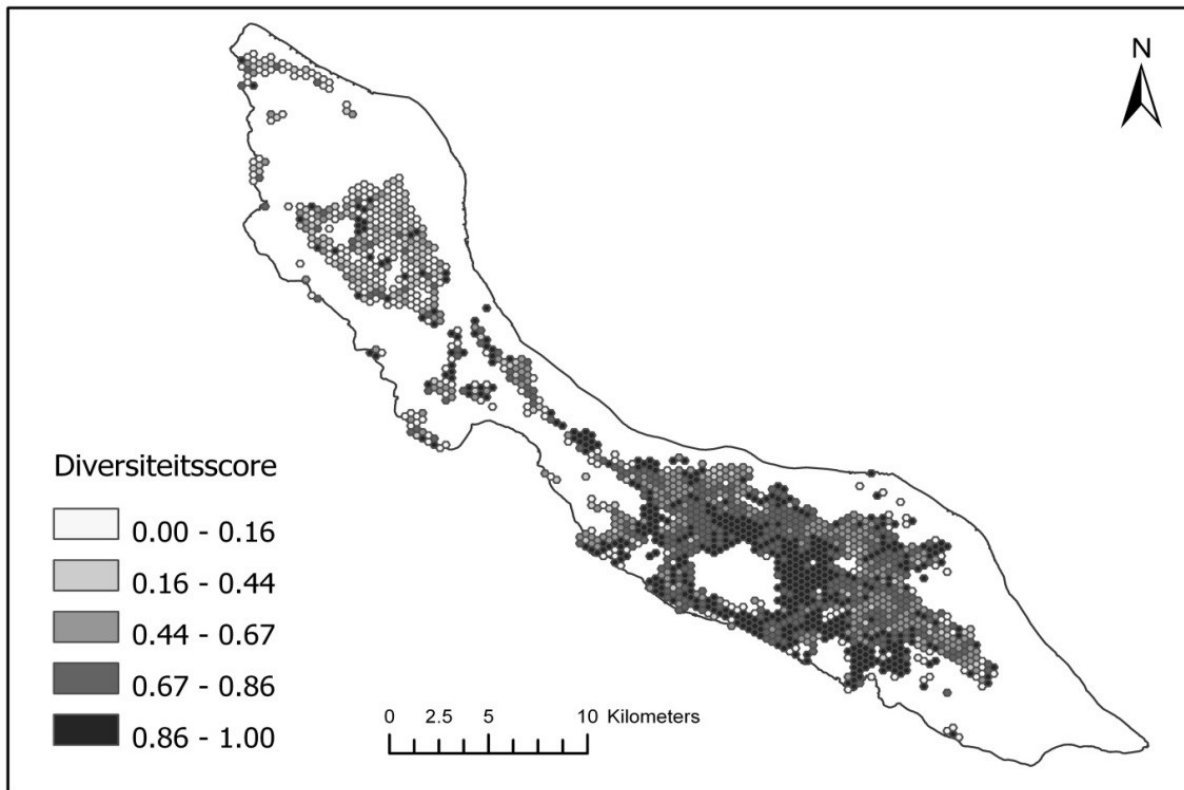
Hier zijn twee verschillende oppervlaktes gebruikt voor de vierkanten en hexagonalen, namelijk 10 hectare en 100 hectare. De ruimtelijke verdeling van de diversiteitsindex is veel duidelijker te zien dan bij de choropleet. Uiteraard is die verdeling gedetailleerder bij de vierkanten en de hexagonalen met het kleinste oppervlakte. Zaken zoals de boodschap die de visualisatie over moet brengen, de aard van de indicator en de ruimtelijke spreiding zijn onder meer bepalend voor de grootte van het oppervlakte van afzonderlijke vierkanten of hexagonalen, oftewel de fijnmazigheid van het raster. Wanneer bijvoorbeeld een globaal beeld verkregen moet worden is een grotere oppervlakte (een grover raster) vaak te verkiezen boven een kleinere oppervlakte (een fijner raster) die lastiger te interpreteren is. Ook bij het optreden van ruimtelijk zeldzame observaties is het verstandig om te werken met een relatief groot oppervlakte. In de praktijk moet deze zaken overwogen worden om tot een zo goed mogelijke visualisatie te komen.

Wanneer de vierkante rasters vergeleken worden met de hexagonale rasters blijkt dat de vierkante rasters vertekend kunnen werken. De buitengrenzen zijn hoekiger dan ze hadden moeten zijn op basis van de onderliggende gegevens. Vierkante rasters zijn in het algemeen minder geschikt om organisch gevormde gebieden of patronen met veel bochten te visualiseren, daarvoor zijn hexagonale rasters beter geschikt. Dit effect wordt in de Engels literatuur een *edge effect* genoemd. Wanneer de choropleet vergeleken wordt met de raster kaarten, valt op dat vooral de dunbevolkte gebieden een vertekening kunnen laten zien. Bij bijvoorbeeld de buurt Oostpunt wordt verwacht dat de diversiteit erg hoog is voor het gehele oppervlakte dat de buurt beslaat. Op basis van de figuren van de rasters, vooral de gene met kleine gebieden van 10 hectare valt op dat er bijna geen datapunten liggen in de buurt Oostpunt. Oostpunt is dan ook dunbevolkt. Er is hier sprake van een vorm van area bias.

**Figuur 4. Diversiteitsindex per hexagonal (100ha), Curaçao, Census 2011**



**Figuur 5. Diversiteitsindex per hexagonaal (10ha), Curaçao, Censur 2011**

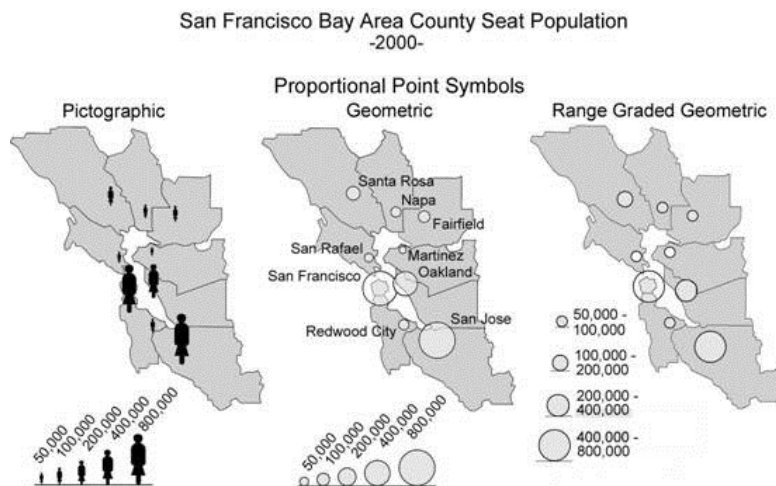


## Andere visualisatiemethoden

Behalve het gebruik van een choropleet en rasters zijn er ook andere visualisatiemethoden bruikbaar. Een paar populaire en bruikbare alternatieven zijn de figuratieve kaarten en dichthedenkaarten. Bij figuratieve kaarten (figuur 6) wordt met symbolen statistische informatie gekoppeld aan een bepaald gebied. Vaak wordt als symbool een cirkel of vierkant gebruikt die in grootte varieert. De grootte van de symbolen wordt bepaald door de waarde van de indicator. Natuurlijk kan er ook gekozen worden voor andere visualisaties zoals staafdiagrammen of boxplots. Een dergelijke kaart, waarin per gebied een staafdiagram, boxplot of vergelijkbaar figuur wordt afgebeeld die de waarde(n) van de indicator weergeeft, wordt een *kartogram* genoemd. In dit laatste geval is het vaak niet verstandig om de grootte van symbolen te laten bepalen door de waarde van de indicator. De interpretatie en de grafiek is dan niet duidelijk af te lezen en te vergelijken met andere gebieden.



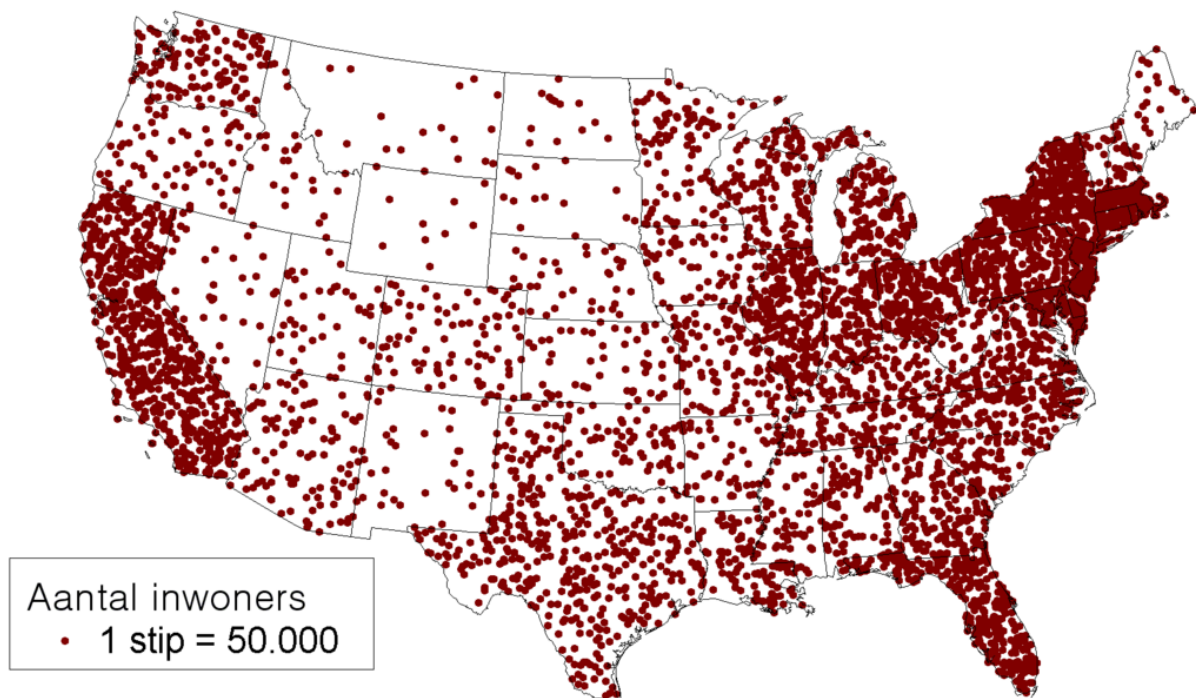
**Figuur 6. Voorbeeld van een figuratieve kaart**



Bron: <http://gis.depaul.edu/shwang/teaching/geog258/QuantitativeThematicMap.htm>

Bij dichthedenkaarten (figuur 7) wordt in een gebied een aantal symbolen geplaatst. Vaak zijn dit stippen. De gebieden worden niet ingekleurd. Het aantal symbolen wordt evenredig met de waarde (of een veelvoud van de waarde) van de indicator in een gebied genomen. Deze kaarten zijn goed te gebruiken als men geïnteresseerd is in spreidingspatronen, zoals de manier waarop mensen zich verspreiden over een land. Groot voordeel van dit type visualisatie is dat grenzen en oppervlaktes geen grote functie hebben in de interpretatie, maar wel gebruikt kunnen worden ter oriëntatie.

**Figuur 7. Voorbeeld van een dichthedenkaart**



Bron: [https://commons.wikimedia.org/wiki/File:Vs\\_stippenkaart.PNG](https://commons.wikimedia.org/wiki/File:Vs_stippenkaart.PNG)

## Mogelijke andere toepassingen GIS

Hoewel in dit artikel gebruik gemaakt is van censusgegevens is het heel goed mogelijk gebruik te maken van steekproefgegevens. Bovendien kunnen ook andere gegevens met een GIS gevisualiseerd worden. Op dit moment vinden er onder meer onderzoeken plaats die zich richten op veldwerk bij het CBS. Daar wordt onderzocht op welke wijze GIS ingezet kan worden om processen tijdens de veldwerkperiode te monitoren en te evalueren. Tevens wordt er onderzocht welke visualisaties zich het beste lenen voor de verwerkingprocessen, de analyse en het publiceren van resultaten.

## Concluderende opmerkingen

Ruimtelijke datavisualisatie kent vele verschillende methoden. Iedere methode heeft zijn voor- en nadelen. Belangrijk is dat de keuze voor een visualisatiemethode afhankelijk is van onder andere de eigenschappen van de te visualiseren data, de boodschap die de kaart dient over te brengen, de methode van publiceren (bijvoorbeeld papieren uitgave of digitale kaart) en het gewenste detailniveau van de visualisatie. De choropleet heeft als belangrijk voordeel dat administratieve grenzen kunnen worden gebruikt, maar onder andere als nadeel dat er *area bias* kan optreden. Het gebruik van rasters heeft als groot voordeel dat alle gebieden in een raster dezelfde vorm en oppervlakte hebben en daardoor eenduidig met elkaar te vergelijken zijn. Een van de nadelen is dat er *edge effects* kunnen optreden die de onderliggende werkelijkheid vertekend weergeven.

Enkele andere ruimtelijke visualisatiemethoden zijn figuratieve kaarten, dichthedenkaarten en kartogrammen.

## Literatuurlijst

- ESRI. (2016, 11 1). *What is GIS?* Retrieved from [www.esri.com/what-is-gis](http://www.esri.com/what-is-gis)
- Magurran, A. (2004). *Measuring Biological Diversity*. Malden, MA: Blackwell Publishing.
- Reardon, S. F., & Firebaugh, G. (2002, January). Measures of multigroup segregation. *Sociological Methodology*, 32(1), 33-67.
- Schmid, C. (1983). *Statistical Graphics, Design Principles and Practices*. New York: John Wiley & Sons.