

AIRBNB OP DE KAART: DE CARTOGRAAF IN HET BIG DATATJDPERK



Het beleidsdebat rondom Airbnb kenmerkt zich door een gebrek aan feitelijke informatie. Ruimtelijke analyse van grote hoeveelheden data biedt hierbij een uitkomst. Bestaande technieken en recente innovaties in GIS en cartografie maken het mogelijk om de ontwikkeling van Airbnb eenvoudig in kaart te brengen. Een kritische blik blijft echter noodzakelijk.

De opkomst van Airbnb kan gerust als stormachtig worden geduid. Van het aanbieden van slaapplekken tijdens een congres in de zomer van 2008 in San Francisco, groeide het platform van Brian Chesky en Joe Gebbia uit tot een platform met meer dan een half miljard boekingen. De destijds gegeven oplossing, een netwerkmoment en een extra zakcentje in het dankzij gentrificatie alsmaar duurder San Francisco, wordt nu juist als een probleem gezien. Airbnb ligt in talloze steden en streken onder vuur omdat het zelf zou zorgen voor gentrificatie. Het is dan ook niet verwonderlijk dat er vanuit verschillende richtingen protest klinkt. Bewonerscollectieven, beleidsmakers, politici, academici en ondernemers mengen zich in het debat en breken zich het hoofd over hoe hier nu mee moet worden omgegaan.

De onzekerheid over de effecten en mogelijke reacties op Airbnb staat echter in contrast tot de mogelijkheden die er zijn voor het analyseren van de ontwikkeling en ruimtelijke spreiding van dit fenomeen. Waar de analyse van de ontwikkeling van toerisme zich in de praktijk vaak beperkt tot het presenteren van overnachtings- en aankomststatistieken, laat Airbnb zich op een

innovatievere manier in kaart brengen. Ondanks dat Airbnb zelf, net als accommodatieaanbieders in de reguliere sector, zeer weinig informatie prijsgeeft, kan tegenwoordig deze informatie uit tweede hand verkregen worden. Voorheen werden analyses binnen het toerisme gebaseerd op de aangifte van de toeristenbelasting of op enquêtes, tegenwoordig kan in het geval van Airbnb informatie door middel van 'scraping' worden verkregen.

Makkelijker vinden en verbinden

Door de opkomst van de platformeconomie is het internet een plaats geworden waar aanbieders en afnemers van producten en diensten elkaar steeds makkelijker kunnen vinden. Via verschillende platformen wordt het delen van producten en diensten niet alleen mogelijk gemaakt, de platformen bieden ook de mogelijkheid om in een vertrouwde en veilige omgeving zaken te doen. Middels 'reviewsystemen' en verzekeringen zijn omgevingen gecreëerd waar tegenwoordig volop gebruik van wordt gemaakt, bijvoorbeeld voor het aanbesteden van klussen (Werkspot), het delen van autoritten (Blablacar) of het in dit artikel besproken aanbieden van

vakantieaccommodaties via Airbnb.

Een belangrijke functionaliteit van een platform is dat het inzicht moet geven in welke aanbieders op welk moment en tegen welke prijs beschikbaar zijn. Niet alleen kunnen potentiële afnemers hierdoor het aanbod filteren, het biedt ook de mogelijkheid voor derden om informatie te verzamelen over het gebruik van het platform. Dit gebeurt middels een techniek die bekend staat als 'scraping': informatie, in dit geval van de webpagina van het platform, wordt handmatig of automatisch verzameld en verwerkt in een database.

In het geval van Airbnb kan op deze manier niet alleen het aantal aanbieders in een gebied nauwkeurig worden verzameld, maar ook bijvoorbeeld het bedrag waarvoor zij hun accommodatie aanbieden, een benadering van de locatie waar ze gelegen zijn, het aantal keer dat ze gereviewed zijn en het profiel van degene die de accommodatie aanbiedt is te vinden op de website. Wanneer dit 'scrapen' systematisch en met een hoge frequentie gebeurt kan er daarnaast ook worden benaderd hoe vaak een Airbnb verhuurd wordt, en hoeveel de aanbieder hiermee verdient. Het systematisch, en vaak automatisch, verzamelen van deze informatie geeft hierdoor een beeld van de aanwezigheid en activiteit van Airbnb-accommodaties in een gekozen gebied.

Airbnb geeft weinig data prijs, maar deze kan uit tweede hand worden verkregen

Momenteel maken verschillende partijen gebruik van deze techniek. Zo zijn er ruimtelijk wetenschappers, activisten en gemeentes die eigen scrapers hebben ontwikkeld met als doel om inzicht te krijgen in de ontwikkeling van Airbnb in verschillende steden. Zo houdt Murray Cox op zijn platform 'inside Airbnb' voor een groot aantal steden, waaronder Amsterdam, Antwerpen en Brussel, bijna maandelijks bij hoe Airbnb zich manifesteert. Ook commerciële partijen houden zich hiermee bezig. Het in de Verenigde Staten gevestigde bedrijf AirDNA heeft een methode ontwikkeld om door middel van zeer frequente bevestigingen van het platform, in combinatie met een eigen algoritme in te schatten hoe vaak een accommodatie verhuurd wordt. Door de uitgebreide scrapingstechnieken worden de data van AirDNA, bij gebrek aan officiële data, momenteel gezien als de meest complete en betrouwbare manier om de ontwikkeling van Airbnb in kaart te brengen, maar aan deze data hangt wel een behoorlijk prijskaartje.

De opkomst van grote data-analyses

Het dagelijks bevragen van een platform, zoals AirDNA dat doet, zorgt voor een enorme hoeveelheid data. Elke Airbnb accommodatie wordt als apart datapunt opgeslagen. Alleen al in Nederland zorgt dit voor een dataset met miljoenen datapunten. De grootte van de dataset brengt uitdagingen met zich mee. Ten eerste is er nood aan generalisatie: miljoenen datapunten die informatie geven over de dagelijkse activiteit van alle ooit actieve Airbnb's in Nederland moeten worden geaggregeerd tot een overzichtelijk bestand waarin voor een bepaalde periode de activiteit en aanvullende informatie per Airbnb wordt weergegeven. Wanneer we dit bekijken voor het jaar 2018, kan de miljoenen regels tellende database worden omgevormd naar een Excelbestand met slechts 65.000 regels: één voor elke actieve Airbnb.

Volgens de data van AirDNA waren er in 2018 in Nederland ruim 65.000 Airbnb's actief. Deze Airbnb's zijn in dat kalenderjaar tenminste één nacht verhuurd. Gemiddeld werden de Airbnb's zo'n 84 nachten verhuurd, waarbij sprake is van een scheve verdeling: een kleine groep, waaronder veel vakantiewoningen, wordt heel vaak verhuurd, gevolgd door een lange staart van meer sporadisch verhuurde Airbnb's. Door via GIS-analyses de Airbnb's te koppelen aan gemeentes kan ook op lokaal schaalniveau een analyse gemaakt worden. Zo blijkt bijvoorbeeld dat in grote steden het percentage overnachtingen in Airbnb's veel hoger is dan het landelijk gemiddelde (zo'n 4,5 procent of 5,5 miljoen overnachtingen): in Rotterdam vonden zo'n 15 procent van alle toeristische overnachtingen plaats in een Airbnb en in Utrecht zelfs bijna een derde.

Hoewel de beschrijvende statistieken al een interessant inzicht geven in de ontwikkeling van Airbnb in Nederland, staan de verzamelde data een veel diepgaandere analyse toe. 65.000 datapunten verspreid over Nederland zijn nog een te grote hoeveelheid om succesvol op een kaart weer te geven. Recentelijk zijn er verschillende methoden ontwikkeld om grote(re) datasets inzichtelijk te kunnen maken. Een in populariteit toenemende methode is de 'heatmap'. In een heatmap worden de datapunten geaggregeerd en weergegeven als felle oplichtende vlekken wanneer er een hoge dichtheid van punten aanwezig is en een egale neutrale kleur wanneer er een lage dichtheid is. De oplichtende vlekken laten zo zien waar er sprake is van een 'verhit gebied', oftewel waar veel Airbnb's op een kleine ruimte bij elkaar liggen.

Een heatmap van Airbnb's in Nederland is makkelijk te maken. Een Excel bestand met alle 65.000 Airbnb's wordt geüpload naar ArcGIS Online, ArcGIS Insights of een open source programma als QGIS. Vervolgens kan deze data via een intuïtief menu worden gevisualiseerd als een heatmap. De enige voorwaarde is dat elke Airbnb is voorzien van X en Y coördinaten. Deze vorm van ruimtelijke analyse is toegankelijk voor een breed publiek en er is zeer weinig cartografische achtergrondkennis nodig om een inzichtelijke visualisering te creëren. Dit brengt echter risico's met zich mee. Met de analyse uit figuur 1 als voorbeeld, een heatmap van alle in 2018 actieve Airbnb's gemaakt in ArcGIS Online, illustreren we hoe een ogenschijnlijk duidelijk resultaat een complexe achtergrond heeft en dat het noodzakelijk is om hier meer van te weten om de juiste analyses en visualisaties te doen.

Figuur 1: 'heatmap' van de verspreiding van actieve Airbnb's in Nederland in 2018 (gemaakt in ArcGIS online met data AirDNA)



Hete steden: het ruimtelijk patroon van Airbnb

Het ruimtelijk patroon in de heatmap in figuur 1 is op het eerste oog erg duidelijk. Enkele steden lichten op: Rotterdam, Den Haag, Nijmegen en Utrecht zijn op deze kaart matig warm, maar bijvoorbeeld Groningen en de omgeving van Maastricht zijn op deze kaart opvallende verhitte plaatsen. Amsterdam is de Airbnb-hotspot bij uitstek, ongeveer de helft van alle Nederlandse Airbnb's zijn in de hoofdstad gelegen, maar ook het gebied van de Zaanstreek tot aan de regio Alkmaar ten noorden van Amsterdam licht op in de kaart. Verder zijn er oplichtende gebieden te vinden rond traditioneel veel bezochte toeristische regio's zoals de Veluwe, in Friesland en op de Waddeneilanden.

De 'heatmap' geeft een inzichtelijke samenvatting van ruim 65.000 datapunten. Ze geeft echter een vertekend beeld weer van de werkelijkheid: de analyse doet geen recht aan de complexiteiten die met het werken met grote geo-data verbonden zijn. Het is bijvoorbeeld voor de gebruiker van de toepassing niet duidelijk op welke manier data geaggregeerd wordt en wat de kleurstelling precies betekent. Op deze manier kunnen de resultaten intuïtieve maar foutieve resultaten weergeven. Maastricht en Groningen vallen in de heatmap bijvoorbeeld enorm op, maar hebben in werkelijkheid maar half zoveel Airbnb's in vergelijking met Utrecht, Rotterdam en Den Haag. Het gebied reikend van de Zaanstreek tot Alkmaar is in de heatmap een Airbnb-hotspot, maar in werkelijkheid bevinden zich er in de Zaanstreek maar ongeveer 500 Airbnb's, en rond Alkmaar zo'n 700 Airbnb's. Dit is vergelijkbaar met het aantal Airbnb's gelegen in het gebied van Eindhoven tot en met 's-Hertogenbosch, welke in de heatmap niet opvalt. De vraag die nu

rest is wanneer de ArcGIS online applicatie een gebied als 'heet' beschouwt. De manier waarop de 'heatmap' wordt gemaakt is hierdoor een zogenaamde 'blackbox': het is niet duidelijk wat er gebeurt tussen het moment dat de data erin worden gestopt en het moment dat de resultaten eruit rollen. Een gebrek aan inzicht in de methodologie van de analyse en visualisatie van de data zorgt dat de kaarten niet goed kunnen worden geïnterpreteerd.

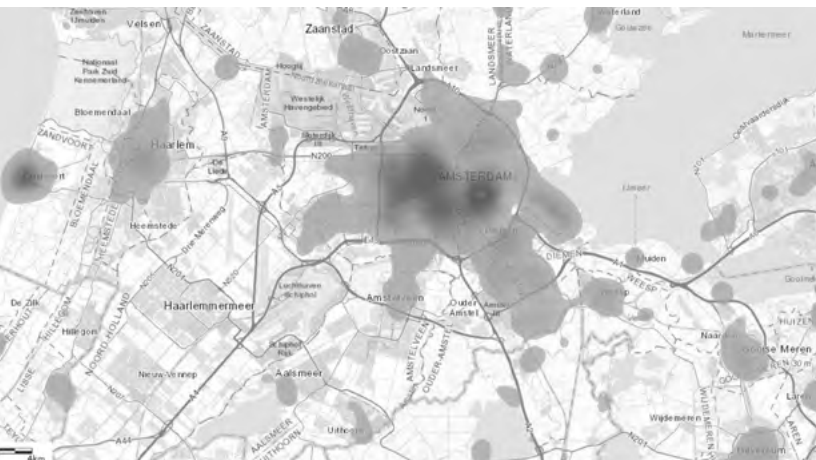
De 'heatmap' is een 'blackbox': het is niet duidelijk wat er gebeurt tussen het moment dat de data erin worden gestopt en het moment dat de resultaten eruit rollen

Op haar website geeft ESRI bij de documentatie over de heatmap-optie aan dat deze slechts dient als 'visuele hulp' en geen nauwkeurige manier is om de verspreiding van een fenomeen in kaart te brengen. Daarnaast stelt ESRI dat deze toepassing bij een groot aantal punten niet meer goed werkt. Hiervoor zou men beter kiezen voor statistische dichtheidsberekeningen zoals de Getis-Ord hotspot analyse. Het valt op dat wanneer er op de heatmap wordt ingezoomd, er andere patronen tevoorschijn komen (figuur 2 & 3). De visualisatie van de heatmaps verandert mee met een aanpassing aan de schaal van de kaart. Hoewel de methode achter de visualisatie 'constant' is, dat wil zeggen dat de hiteschaal niet verandert wanneer in- of uitgezoomd wordt, wordt de dichtheid wel berekend aan de hand van de data die op dat moment in het kaartbeeld zichtbaar is. 'Hete' gebieden zijn heet in relatie tot de gemiddelde waarde in het kaartbeeld. De 'hotspot' die in figuur 1 rondom Amsterdam zichtbaar is, blijkt in figuur 2 vooral geconcentreerd te zijn in het oostelijk gedeelte van de historische binnenstad en de buurten ten westen van de binnenstad, maar binnen de ring A10.

In Zuid-Holland (figuur 3) is het patroon nog extremer. De binnensteden van Den Haag, Rotterdam, Delft en de badplaats Scheveningen lichten iets op, terwijl de rest van de provincie nauwelijks Airbnb's lijkt te hebben. Dit beeld contrasteert enorm met de werkelijkheid, maar wordt veroorzaakt doordat de hiteschaal bepaald is op basis van de data van heel Nederland, en de hoge dichtheid aan Airbnb's in Amsterdam ervoor zorgt dat de rest van Nederland wat verbleekt. Het vergelijken van figuur 2 en 3 zegt dus vooral dat de aanwezigheid van Airbnb's in de regio Den Haag-Rotterdam minder sterk ruimtelijk geclusterd is dan in de regio Amsterdam: het zegt weinig over de absolute aanwezigheid van Airbnb's.

De positie van Amsterdam en de 'constante' methodologie die de schaal van de visualisatie bepaalt verklaart echter nog niet waarom in figuur 1 verschillende plaatsen onverwacht sterk oplichten. De manier waarop de applicatie berekent of er sprake is van een hoge dichtheid verklaart dit echter wel. Er wordt gebruik gemaakt van een algoritme op basis van de 'Kernel' benadering. In deze benadering wordt het gehele onderzoeksgebied in een raster verdeeld, en voor elke cel in het raster wordt er zowel geteld hoeveel punten er in de cel vallen als hoeveel punten er in de directe omgeving liggen. Door middel van een zoekafstand bepaalt het algoritme welke cellen worden meegeteld, en daarbij worden de nabijgelegen

Figuur 2: heatmap van de verspreiding van Airbnb's in de regio rondom Amsterdam in 2018 (gemaakt in ArcGIS online met data AirDNA)



Figuur 3: heatmap van de verspreiding van Airbnb's in de regio rondom Den Haag-Rotterdam in 2018 (gemaakt in ArcGIS online met data AirDNA)



cellen zwaarder geteld dan de cellen die verder weg liggen. Er is sprake van een 'heet gebied' wanneer een cel met een hoge waarde (veel Airbnb's) direct wordt omgeven door andere cellen met hoge waarden.

Hoewel het niet geheel duidelijk is hoe het algoritme wordt toegepast in de 'heatmap', welke zoekafstand wordt gebruikt en hoe groot de rastercellen zijn, lijken de positie van Groningen en Maastricht op te vallen doordat ze een relatief kleine stedelijke kern hebben met redelijk veel Airbnb's, omgeven door een groot gebied met weinig tot geen Airbnb's (in Maastricht komt dit ook door de ligging in het grensgebied), waar bij andere steden, zoals Rotterdam, Den Haag en Utrecht de overgang tussen gebieden met veel en weinig Airbnb's veel geleidelijker is. Dit is een andere conclusie dan wat men kan vermoeden wanneer men de kaarten zonder context bekijkt.

Inzicht in de Blackbox

In een digitale wereld waar grote hoeveelheden data worden verzameld en softwareontwikkelaars met intuïtieve mogelijkheden komen om deze data te analyseren en visualiseren is er meer behoefte dan ooit aan context. De analyse van de heatmap-toepassing laat zien dat deze geen duidelijk beeld geeft van de spreiding van Airbnb's over Nederland. De 'blackbox' waarin de analyse plaats vindt die de visualisatie mogelijk maakt, maakt het de gebruiker makkelijk om snel een analyse van grote databestanden uit te voeren en deze op een aantrekkelijke manier te karteren. Het zorgt er echter ook voor dat er op de achtergrond allerlei keuzen worden gemaakt die invloed hebben op de presentatie waar de gebruiker geen zicht op heeft. In deze context is het belangrijk dat de hedendaagse cartograaf en/of ruimtelijk wetenschapper in staat moet zijn om zelf aan de knoppen te kunnen draaien, en zo inzicht te krijgen en te geven in de 'blackbox'.

In een digitale wereld waar grote hoeveelheden data worden verzameld, is er meer behoefte dan ooit aan context

Wanneer we de ruimtelijke spreiding van Airbnb's in de provincie Zuid-Holland bijvoorbeeld analyseren met de methode van Getis-Ord zien we dat er significante ruimtelijke clusters van Airbnb zijn in de grote steden zoals Den Haag en Rotterdam, in historische steden als Delft en Leiden, maar ook in traditionele toeristische bestemmingen langs de kust. De Getis-Ord methode bepaalt of een fenomeen willekeurig over de ruimte verspreid is, of er sprake is van een significante hotspot, dan wel of er ruimtelijke autocorrelatie optreedt. Een donkergekleurde cel staat voor een gebied met relatief een groot aantal Airbnb's, omgeven door andere gebieden met een groot aantal Airbnb's. Het patroon eindigt abrupt en gaat vrijwel overal direct over in een coldspot. De hot- en coldspots zijn significant, dus met een zekerheidsmarge van 95% kan gesteld worden dat dit patroon niet op toeval rust. Wanneer deze analyse, die ook in programma's als ArcGIS en QGIS beschikbaar is, uitgevoerd wordt, worden alle relevante keuzes en statistieken gegeven. De analist kan hierdoor zien wat er achter de schermen gebeurt, en waar nodig aan de knoppen draaien. De resultaten van de analyse, zeker wanneer deze door de tijd worden gevolgd, kunnen belangrijke inzichten geven over de ontwikkeling van een fenomeen als Airbnb. Ze laten zien in welke plaatsen de

dichtheid van Airbnb's toeneemt, waar nieuwe hotspots beginnen te ontstaan en waar ze ontbreken terwijl ze misschien wel verwacht worden.

In Rotterdam vond zo'n 15 procent van alle toeristische overnachtingen plaats in een Airbnb

De opkomst van inzichtelijke en intuïtieve applicaties, in combinatie met een sterk groeiende beschikbaarheid van databronnen, hebben het voor een breed publiek mogelijk gemaakt om zelf relatief complexe ruimtelijke analyses en visualisaties te maken. Voor cartografen en ruimtelijk wetenschappers is het echter de taak om kritisch te blijven kijken naar deze analyses, want het is, bedoeld of onbedoeld, nog nooit zo makkelijk geweest om te liegen met kaarten.

Literatuurselectie

Getis, A., & J.K. Ord (2010) The analysis of spatial association by use of distance statistics. In Perspectives on Spatial Data Analysis. Heidelberg: Springer, pp. 127-145.

Ioannides, D., M. Röslmaier & E. van der Zee (2018) Airbnb as an instigator of 'tourism bubble' expansion in Utrecht's Lombok neighbourhood. Tourism Geographies, pp. 1-19.

van der Zee, E., D. Bertocchi & D. Vanneste (2018) Distribution of tourists within urban heritage destinations: a hot spot/cold spot analysis of TripAdvisor data as support for destination management. Current Issues in Tourism, pp. 1-22.

Egbert van der Zee (e.l.vanderzee@uu.nl) is universitair docent sociale geografie en planologie aan de universiteit van Utrecht. Hij houdt zich bezig met thema's als de impact van toerisme op steden en gebruikt hierbij GIS in onderwijs en onderzoek. **Demi van Weerdenburg** (dvanweerdenburg@gmail.com) is hoofdredacteur van AGORA en adviseur bij Areaal Advies en gebruikt GIS in de beleidspraktijk en in haar onderzoek naar toerisme en recreatie in steden.

Figuur 4: Getis-Ord hotspot analyse van Airbnb's in de provincie Zuid-Holland in 2019.

