# Leeswijzer en toelichting

Onderzoek ‘Kwetsbare groepen in beeld’

## Kwetsbaarheid

Kwetsbare inwoners zoals gedefinieerd in dit onderzoek verschillen van de overige inwoners op alle aspecten van de gezondheid. Ze zijn vaker eenzaam, hebben een hoger risico op angst en depressie, ervaren vaker onvoldoende regie over hun eigen leven en beoordelen hun eigen gezondheid minder goed. Zij voelen zich vaker beperkt door een chronische aandoening, vanwege hun gehoor, zicht of vanwege verminderde mobiliteit. Ze zijn vaker vermoeid zonder duidelijke reden. Kennen een ongezonde leefstijl. Tot slot hebben ze te maken met een relatief hoge samenloop van beperkingen

Het concept ‘kwetsbaarheid’ is in dit onderzoek opgebouwd uit 5 bouwblokken[[1]](#footnote-1) met daarachter 27 vraagstellingen (of features) uit de GGD Gezondheidsmonitor[[2]](#footnote-2) die gaan over de beleving van gezondheid en beperkingen in de brede zin van het woord.

Het doel van de analyse is om -*op een ongestuurde wijze*- verschijningsvormen van kwetsbaarheid te herleiden, en van elkaar te onderscheiden op basis van het belevingspatroon van *groepen* inwoners.

Met dit onderzoek willen de betrokkenen[[3]](#footnote-3) laten zien dat je met reeds beschikbare dataregistraties aan de hand van vernieuwde methoden en techniek van onderzoek andersoortige inzichten kunt neerzetten. Inzichten die ondersteunend zijn voor beleid en uitvoering.

## Doelgroep

Uit de totale netto-respons van de GGD Gezondheidsmonitor[[4]](#footnote-4) is aan de hand van vijf inclusiecriteria[[5]](#footnote-5) de groep (potentieel) kwetsbare inwoners geïsoleerd:

Eenzaamheid: (zeer) ernstig eenzaam (GGEES203 >7)

Regie op het leven: onvoldoende regie over eigen leven (GGRLS202<24)

Angst of depressie: matig of hoog risico op angststoornis of depressie (GGADS201>21)

Gezondheidsbeleving: (zeer) slechte (ervaren) gezondheid (KLGGA207==3)

Multiproblematiek: hoge mate van samenloop van beperkingen (op de features) (score\_zw>4)

Een inwoner wordt toegerekend aan de groep (potentieel) kwetsbaren als minimaal aan een van de criteria is voldaan.

De onder-/bovengrens voor respectievelijk de criteria ‘samenloop van beperkingen’, ‘gebrek aan regie’, ‘angst & depressie’ en ‘eenzaamheid’ is op een exploratieve wijze bepaald. Hierbij is o.m. afgeleid bij welk niveau van samenloop van beperkingen de regie op het leven significant afneemt.

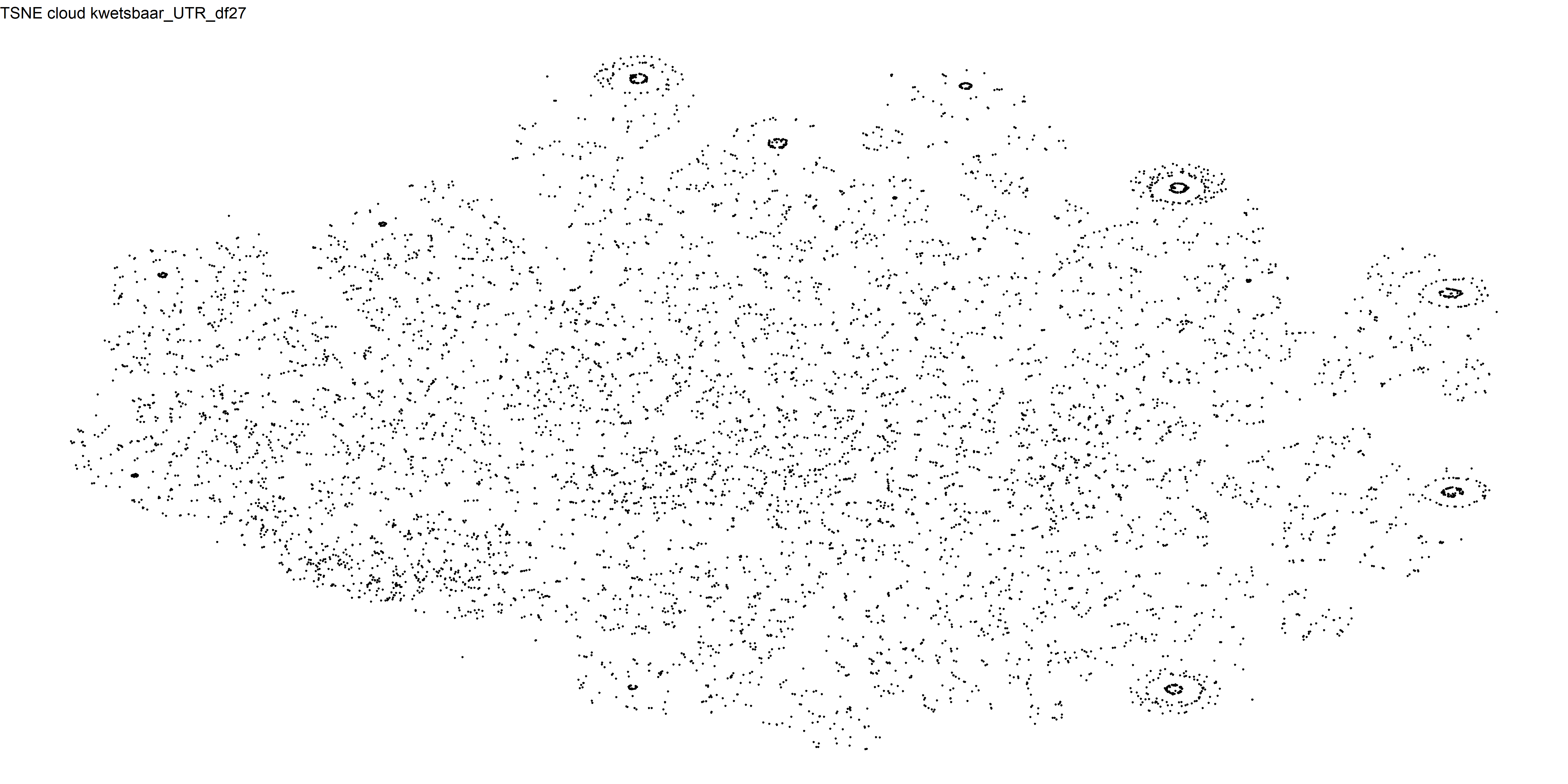
## Dimensionaliteitsreductie

Dimensionaliteitsreductie is het proces om de dimensies van gegevens (i.c. de 27 features) te verminderen waarbij zoveel mogelijk van de informatiewaarde blijft behouden. Dimensionaliteitsreductie ligt aan de basis van het visualiseren van hoog-dimensionale data (de features). Als een dataset meer dan drie dimensies heeft, wordt het moeilijk om deze op een zodanige manier te visualiseren die voor de mens makkelijk te interpreteren is. Daarom wordt een dimensionaliteitsreductie methode gebruikt om de data te reduceren naar een laag-dimensionale representatie. In dit geval tot twee dimensies.

T-Distributed Stochastic Neighbor Embedding (t-SNE) is een algoritme voor dimensionaliteitsreductie die bijzonder geschikt is voor de visualisatie van hoog-dimensionale gegevenssets. Het algoritme streeft naar het behoud van de afstand tussen punten uit de hoger dimensionale ruimte, waarmee punten die dicht bij elkaar liggen verwijzen naar vergelijkbare beleving.

In onderstaande figuur zijn alle kwetsbare inwoners weergegeven uit de provincie Utrecht.

Fig. 1 Kwetsbaren op de radar (dimensionaliteitreductie)



In de figuur zijn verdichtingen of concentraties zichtbaar, maar nauwelijks is te bepalen welke overeenkomstige situaties van kwetsbaarheid er zijn. Daarnaast is het onduidelijk met hoeveel verschijningsvormen we hier te maken (kunnen) hebben.

## Clustering

De eerstvolgende stap is om te bepalen hoeveel verschijningsvormen er idealiter wenselijk zijn. Dit bepalen we aan de hand van de Silhouette methode. De Silhouette waarde is een maat voor hoe vergelijkbaar een observatie is met zijn eigen cluster (cohesie) in vergelijking met andere clusters (differentiatie). Voor de provincie Utrecht zou een aantal van 3 tot 7 verschijningsvormen de situatie het best beschrijven. We kiezen hier voor 7 verschijningsvormen.

De keuze van clusteringsmethode kan een invloed hebben op hoe goed een bepaalde clustering zal scoren qua lading, cohesie en differentiatie. Daarom maken we gebruik van verschillende methoden van clustering: K-means en K-mediods. Achteraf toetsen we welke methode het beste resultaat geeft.

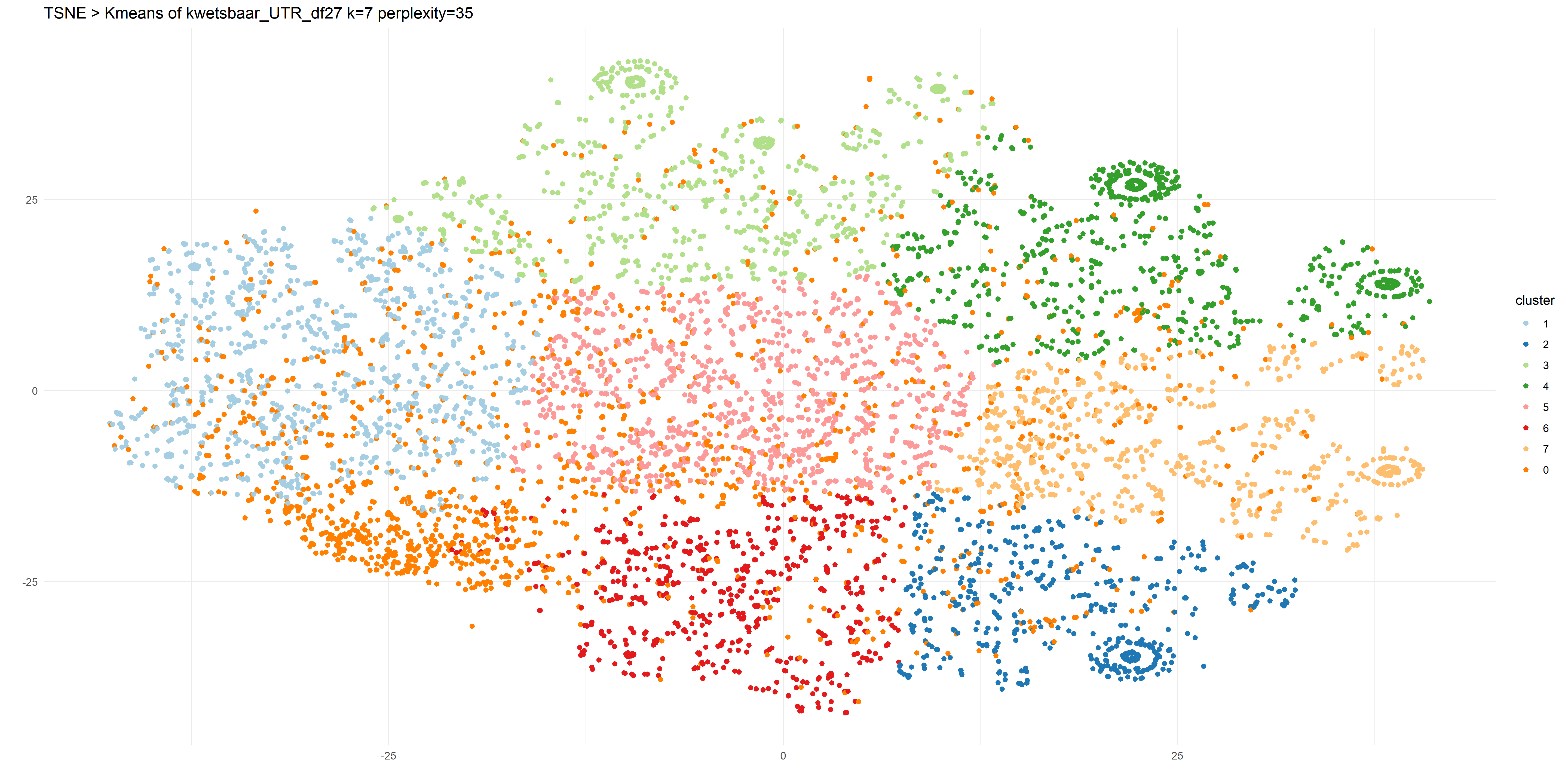
De eerste methode om te komen tot samenhangende en onderscheidende verschijningsvormen maakt gebruik van het K-means clustering algoritme. K-means streeft naar het beperken van de afstand van een observatie (kwetsbare inwoner) tot het centrale punt in een cluster. Dit centrale punt is niet-noodzakelijkerwijs een daadwerkelijke respondent. Het gemiddelde punt dient als het centrum van een cluster. Het K-means clustering-algoritme is daarmee gevoelig voor uitbijters, omdat een gemiddelde gemakkelijk wordt beïnvloed door extreme waarden.

Om toch een stabiele en onderscheidende indeling te verkrijgen worden voorafgaand aan de K-means analyse de extreme uitbijters (in het antwoordpatroon) herleid en geïsoleerd. Er is geen ‘harde’ grens voor wat wel of niet een uitbijter is. Zolang K-means nog samenhangende en onderscheidende clusters tot stand kan brengen voegen we groepen (niveaus) van uitbijters toe.

In onderstaande figuur is een indeling van de groep kwetsbare inwoners met zeven verschijningsvormen geconstrueerd, en zijn de overgebleven uitbijters gemarkeerd met ‘0’ (donker oranje). Uitbijters komen overal voor, met name aan de linkerkant, en meer specifiek linksonder. Let op: er is hier geen sprake van een uniform cluster, of überhaupt een cluster in de zienswijze van de gehanteerde clustering algoritmen.

De figuur maakt inzichtelijk welke personen in een overeenkomstige situatie verkeren. Personen uit een groep met dezelfde kleur (of nummer) kennen in hoge mate een overeenkomstige ervaring en -beperkingen. De observaties nabij een naburig cluster vertonen daarnaast in zekere mate overeenkomsten met het buurcluster.

Fig. 2 Samenhang en verschillen (clustering)



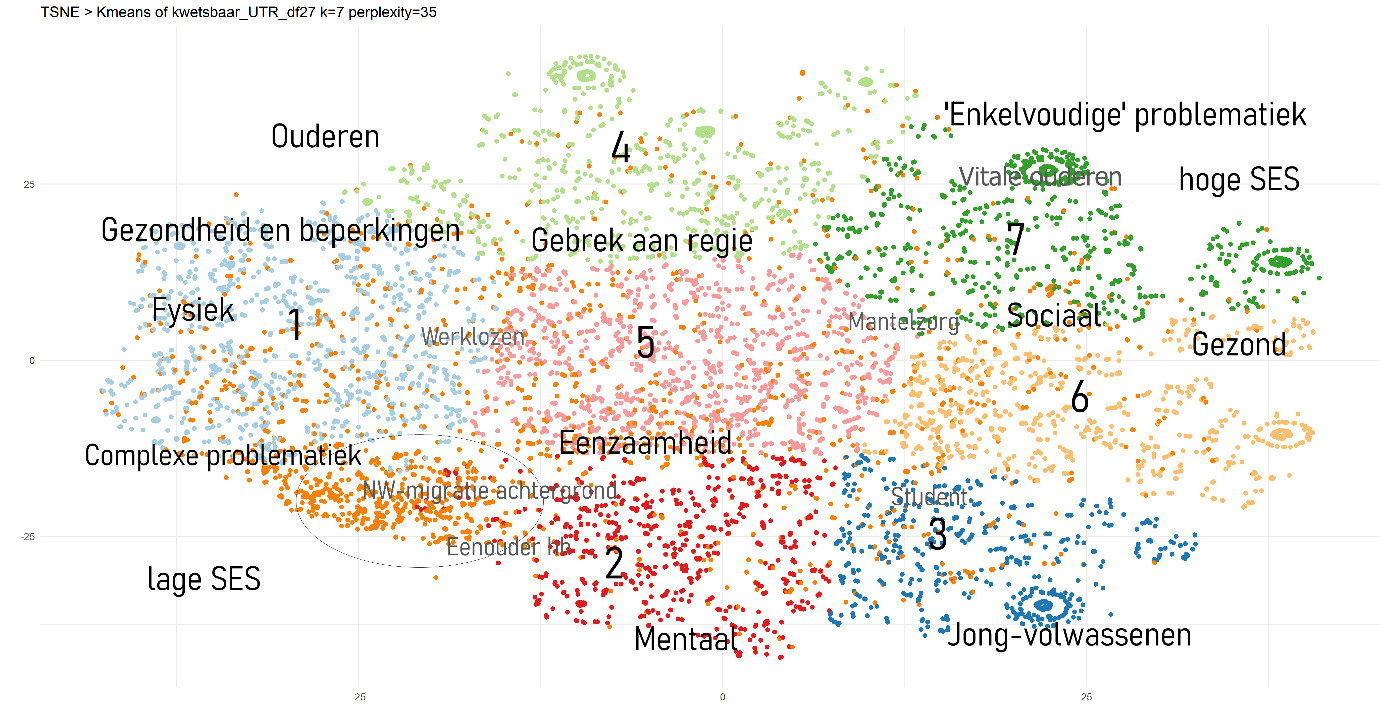
Op basis van de achterliggende kenmerken (w.o. de features) van de clusters kan worden opgemaakt waar de deelgebieden van de radarplot betrekking op hebben[[6]](#footnote-6). Voor dit moment zijn we alleen geïnteresseerd in de overeenkomstigheid en verschillen van *meerdere* clusters. Deze zogenaamde latente dimensies helpen bij de globale interpretatie van clusters en de onderlinge verhouding.

In principe zijn de x-as en de y-as niet eenduidig te definiëren, maar er zijn in dit geval wel dimensies die (nagenoeg) parallel lopen met deze assen. Op de x-as is bv. de sociaaleconomische status (SES) afgetekend. Op de x-as zijn ook de drie dominante vormen van beperkingen te herkennen (fysiek, mentaal, sociaal). Personen aan de linkerkant van de plot zijn hoog-kwetsbaar, en personen aan de rechterkant in lage mate kwetsbaar. Aan de linkerkant kennen personen een slechte ervaren gezondheid, aan de rechterkant is de ervaren gezondheid overwegend goed.

Op de y-as beweegt zich leeftijd (of levensfase). Aan de bovenkant van plot bevinden zich personen zonder betaalde baan en/of met een beperkt zinvolle dagbesteding.

Op de diagonaal van linksonder naar rechtsboven tekent de complexiteit van de problematiek zich af. Je kunt er ook een dimensie vitaliteit (of weerbaarheid) in lezen. Vrijwel parallel daaraan (maar net er onder) loopt de dimensie eenzaamheid.

Fig. 3 Latente dimensies en dominante kenmerken



Met deze redenatie kun je ook de snijpunten van de latente dimensies verder duiden. Links onder bevinden zich personen met een hoge mate van samenloop van beperkingen (‘multiproblematiek’), ze onderscheiden zich van de andere verschijningsvormen doordat fysieke uitdagingen een dominante component vormen. De personen zijn (jong-)volwassen en kennen een lage sociaaleconomische status. Eenzaamheid speelt hierbij een relatief grote rol.

Iets hoger in de radarplot bevinden zich personen van hogere leeftijd, eveneens met een slechte (ervaren) gezondheid, en met beperkingen die hoofdzakelijk gerelateerd zijn aan het fysiek functioneren. Op hogere leeftijd krijgen deze ouderen ook te maken met gebrek aan regie op het eigen leven.

|  |
| --- |
| Het is van belang op te merken dat de clusters zich vooral onderscheiden in een *dominante* component in de beleving van gezondheid en beperkingen. Cluster 1 onderscheidt zich vooral door de fysieke beperkingen, terwijl ook sociale en mentale beperkingen in zeker mate opspelen. In hoofdzaak hebben de personen te maken met een lage sociaaleconomische status, maar het is denkbaar dat er bijvoorbeeld personen met een hoge opleiding in voorkomen. Dit heeft te maken met het feit dat de bouwblokken van het concept kwetsbaarheid beperkingen en uitdagingen (‘belevingsperspectief’) weerspiegelen en geen persoonskenmerken zoals opleiding of leeftijd omvatten. De latente dimensies SES en leeftijd volgen (indirect) uit de correlatie met features. |

Observaties in een grensgebied van clusters vertonen naast het dominante karakter van het eigen cluster ook in zekere mate overeenkomsten met een of meerdere buurcluster(s). In het grensgebied van cluster 1 en 2 (het bereik van de meer zorgwekkende situaties) loopt fysieke problematiek samen met mentale uitdagingen (angst en depressie) en komt ook eenzaamheid om de hoek kijken. Het is niet voor niets dat in dit gebied veel bijzondere situaties (extremen) voorkomen. Dit zijn de complexe multiproblematiek situaties. Hieronder vallen onder meer personen met een niet-Westerse migratieachtergrond en eenouderhuishoudens met een laag inkomen.

De hardnekkige concentratie van de outliers (zorgwekkende inwoners) in het voornoemde grensgebied is kenmerkend voor de provincie Utrecht. Ze zijn in hoge mate bepalend voor het zorgwekkende karakter van cluster 1 en 2.

Doordat de extreme observaties (uitbijters) tot op heden buiten de indeling naar verschijningsvormen is gelaten, blijft een groot deel van de lading -het elementaire karakter- van de clusters onderbelicht. Vooral in de clusters aan de linkerkant van de radarplot.

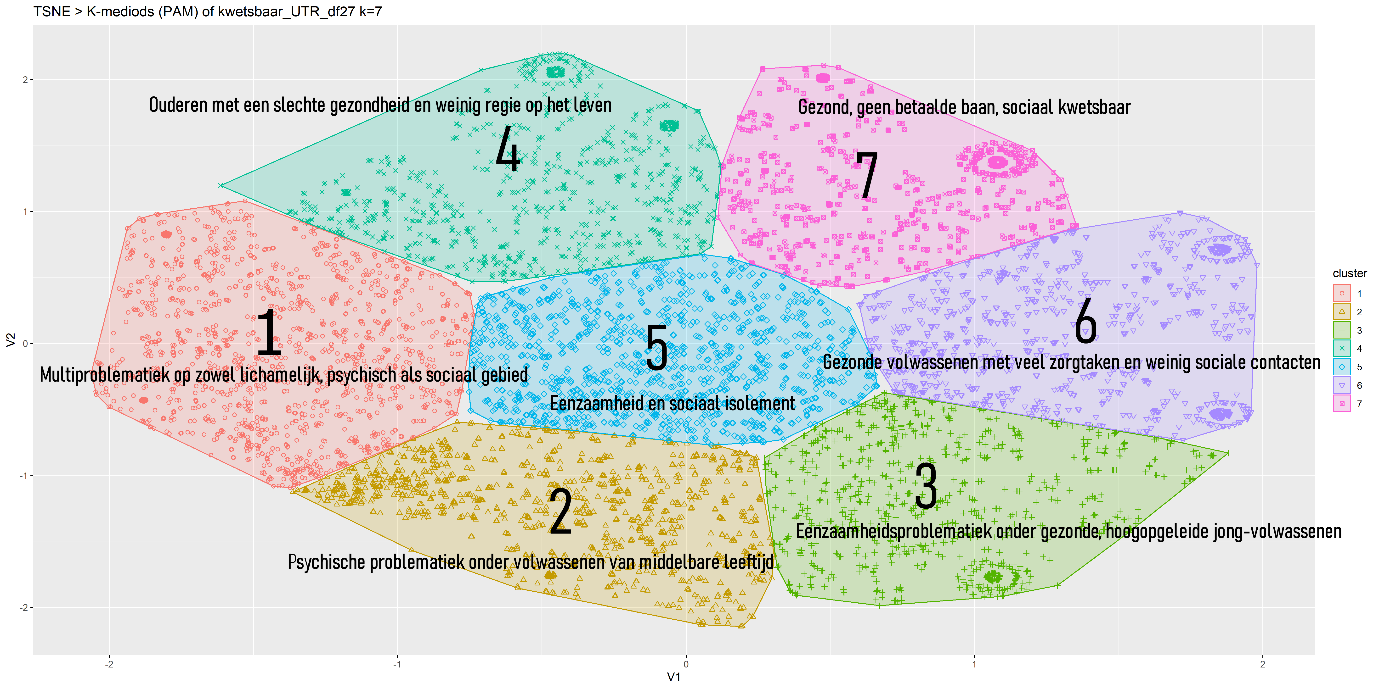
Om toch alle observaties te kunnen verdelen over de zeven verschijningsvormen en (daarmee) meer onderscheid tussen de clusters tot stand te brengen is een alternatieve clustering methode gehanteerd: K-mediods (PAM).

K-mediods streeft er eveneens naar om de ongelijkheid tussen punten in een cluster te beperken. K-mediods gebruikt een feitelijke observatie (kwetsbare inwoner) om een cluster te construeren. Een mediod is de meest centraal gelegen observatie van het cluster, met minimale som van de afstanden tot andere observaties. PAM is in staat om ook bij datasets met een aanzienlijke groep outliers (‘ruis’) een stabiele clustering tot stand te brengen.

Figuur 4 toont de indeling naar de verschijningsvormen van kwetsbaarheid op basis van dit clustering algoritme. De indeling volgt in hoge mate de indeling van de vorige methode. Echter, er zijn wel enkele verschillen. Cluster 6 en vooral cluster 7 liggen hoger op de y-as (gaat nu dus meer over uitdagingen gerelateerd aan hoge(re) leeftijd). Cluster 7 ligt nu ook iets meer naar links (lagere SES, minder regie op het leven, etc.). Het meest opmerkelijke verschil is dat cluster 2 zich nu verder uitstrekt naar links en daarmee veel van de oorspronkelijke outliers uit het grensgebied met cluster 1 een plekje kan geven. Een groot deel van deze complexe en zorgwekkende situaties blijkt daarmee gerelateerd aan uitdagingen op mentaal vlak (onzekerheid, angst en depressie).

Onderstaande clusterindeling is uiteindelijk uitgewerkt en beschreven.

Fig. 4 Uiteindelijke verschijningsvormen van kwetsbaarheid



De K-mediods geeft een relatief hoge mate van samenhang van de observaties in de verschijningsvormen ten opzichte van K-means. De K-means methode is niet goed in staat om de extreme outliers te accommoderen. K-means kan daarmee wel goed bepalen wat de sterk afwijkende situaties zijn per cluster. Alle cases die wel zijn toegewezen aan een K-mediods cluster (cl\_pam), maar niet aan een K-means cluster (cl\_kmeans) vragen om een specifieke interventie of menselijke maat. Het dominante aspect van de kwetsbaarheid wordt geduid door de K-mediods kwalificatie.

## Robuustheid en flexibiliteit

Doordat (o.m.) dimensiereductie- en clustertechnieken zijn gestapeld, ontstaat er een robuuste methode om de verschijningsvormen inzichtelijk te maken. Zolang de bouwblokken in stand blijven zal de uitkomst in hoge mate gelijk zijn aan deze uitkomst. Het is mogelijk om een of enkele elementen (vraagstellingen uit de monitor) binnen een bouwblok aan te passen. Dat heeft nauwelijks invloed op deze uitkomst.

Het is denkbaar dat de bouwblokken worden uitgebreid met een nieuwe dimensies (b.v. vitaliteit en veerkracht, of levenszinvolheid en geluk). Dat is geen enkel probleem zolang het ook hier gaat om ervaren beperkingen en vitaliteit, en het aantal vragen in verhouding staat tot de overige bouwblokken.

## Classificatiemodel

Vanwege het stabiele karakter van de uitkomst is ook een classificatiemodel ontwikkeld. Deze wordt getraind op basis van een grote populatie (meerdere regio’s, Nederland). Het classificatiemodel maakt gebruik van de ruwe of onbewerkte features (de letterlijke antwoorden op de vragen), kan omgaan met missing values in het antwoordpatroon van een respondent, en ondervindt nauwelijks hinder van (bovenmatige) ruis. Het model classificeert elke nieuwe casus (inwoner) op basis van de leerervaring en wijst deze toe aan een verschijningsvorm indien aan de orde[[7]](#footnote-7).

## Wat kun je er mee?

De analyse resulteert in een **foto** van de verschijningsvormen van kwetsbaarheid waarbij de onderlinge verhouding van kwetsbare inwoners en van de verschijningsvormen zich duidelijk aftekent. Het is een aanknopingspunt om een samenhangend en integraal beeld te vormen van de kwetsbare inwoners.

Diverse minder tastbare en **milde(re) vormen** van kwetsbaarheid zijn in beeld gebracht.

Het geeft inzicht in de *daadwerkelijke* **samenloop van beperkingen** en de *relatieve* ‘zwaarte’ van de problematiek.

De verschijningsvormen zijn **gekwantificeerd** naar gemeente en wijk. De opbouw per gebied kan worden vergeleken. Het profiel van de provincie Utrecht kan worden vergeleken met andere (GGD-) regio’s.

De foto vormt een gezamenlijk discussiepunt om met de verschillende rollen en verantwoordelijkheden (de organisatie en stakeholders) om tafel te gaan zitten. De opzet van het onderzoek en de definitie van het begrip kwetsbaarheid zorgt voor een **gemeenschappelijke focus** **en taal**.

Per verschijningsvorm van kwetsbaarheid zijn **typeringen** van de persoons- en situationele kenmerken geconstrueerd (persona’s). Deze vormen een aanknopingspunt om doelgroepen te herleiden en deze te relateren aan de positie binnen de radarplot. Persona’s die in meerdere clusters opduiken kunnen meer doelgericht worden bediend met specifieke of gebundelde interventies. Vb.: er zijn mantelzorgers die in meerdere clusters voorkomen, maar die wel aanzienlijk verschillen in urgentie en overbelasting (afnemen gezondheid, regie eigen leven).

Het biedt een gelegenheid om andere kenmerken te **segmenteren** op basis van deze typologie.

Dit belevingsperspectief kan dienen als **spiegel** voor objectieve inzichten.

Het bestaande **interventieporfolio** kan tegen het licht worden gehouden om te bepalen of het porfolio dekkend is voor de materie en de specifieke situaties en omstandigheden daarbinnen.

De verschijningsvormen van dit onderzoek (2016) kunnen worden **geprojecteerd** op een nieuwe meting aan de hand van een getraind classificatiemodel. Daarmee is er een nulmeting en worden ontwikkelingen inzichtelijk gemaakt. Het is interessant om te volgen of de incidentie per verschijningsvorm toe- of afneemt en hoe de verschijningsvormen zich bewegen over de latente dimensies.

Het algoritme is **vrij beschikbaar** via GitHub. Het mag worden gebruikt voor eigen doeleinden (regio, doelgroepen, andere vraagstukken etc.). Verschillende partijen kunnen bijdragen aan het doorontwikkelen van het algoritme.

https://github.com/VNG-Realisatie/Data-Science-Hub-SES

1. Eenzaamheid, Regie op het leven, Angst en depressie, Gezondheidsbeleving, Multiproblematiek (samenloop van beperking) [↑](#footnote-ref-1)
2. GGD Gezondheidsmonitor volwassenen en ouderen, Provincie Utrecht 2016 [↑](#footnote-ref-2)
3. Uden, Brummen, Westland, GGD Hart voor Brabant en VNG Realisatie/Data Science Hub [↑](#footnote-ref-3)
4. GGD Gezondheidsmonitor volwassenen en ouderen, Provincie Utrecht 2016 [↑](#footnote-ref-4)
5. De drempelwaarden van de criteria volgen uit de procedure en kunnen afwijken per GGD-regio [↑](#footnote-ref-5)
6. zie beschrijving van de verschijningsvormen, en de tabellenbijlage [↑](#footnote-ref-6)
7. Met een betrouwbaarheid van ruim 90%. Daar waar er sprake is van ‘misclassificatie’ gaat het veelal om observaties die zich in een (complex) grensgebied bevinden. [↑](#footnote-ref-7)