# Kwetsbare inwoners in de GGD-regio Utrecht

## Onderzoeksverslag, versie 1

## Den Haag, 2019-11-19

Kwetsbare inwoners zoals gedefinieerd in dit project verschillen van de overige in 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 uitdagingen.

### Doel

* Tastbaar maken van (dreigende) zorgwekkende situaties vanuit belevings- en outcome-perspectief.
* Herleiden, definiëren en kwantificeren van de verschillende verschijningsvormen van kwetsbaarheid

### Scope

De GGD Gezondheidsmonitor (meting 2016) omvat 37.495 respondenten (zelfstandig wonende inwoners van 19 jaar en ouder) uit de GGD-regio Utrecht.

### Operationalisering

#### Dimensies

Het begrip kwetsbaarheid is gedefinieerd op basis van 5 (outcome)dimensies:

Eenzaamheid: (zeer) ernstig eenzaam

Regie op het leven: onvoldoende regie over eigen leven

Angst of depressie: matig of hoog risico op angststoornis of depressie

Gezondheidsbeleving: (zeer) slechte gezondheid

Multiproblematiek: hoge mate van samenloop van uitdagingen

#### Features

De dimensies zijn geladen met een evenwichtige set van 27 features. De features vertegenwoordigen het concept kwetsbaarheid op de belangrijke elementaire aspecten[[1]](#footnote-1). De features zijn dichotoom gecodeerd, waarbij de kwetsbaarheid is vervat in waarde ‘1’ voor het aantwoordbereik dat in hoge mate problematiek of uitdaging vertegenwoordigd.

Zinvol bestaan / betekenisvol sociaal netwerk

MMWSA205, MMWSA211 : werk / opleiding

MMVWB201 : vrijwilligerswerk

MCMZGS203 : mantelzorger

Eenzaamheid

GGEEB201 : !Kan praten over dagelijkse problemen

GGEEB203 : Ervaar leegte

GGEEB204 : !mensen om op terug te vallen bij narigheid

GGEEB207 : !Veel mensen om op te vertrouwen

GGEEB208 : !Voldoende mensen waarmee verbondenheid is

Gezondheid en beperkingen

CALGA260 : Heeft langdurige ziekte(n) of aandoening(en)

CALGA261 : Is (ernstig) beperkt in activiteiten vanwege gezondheid

LGBPS203, LGBPS204, LGBPS205 : Beperking met horen, zien of mobiliteit

MMIKB201 : Moeite met rondkomen

Lifestyle / gedrag

AGGWS205 : Obesitas, ofwel een BMI van 30 of hoger

LFALA213 : Zware drinker

LFRKA205 : Roker

Angst en depressie

GGADB201 : Vaak vermoeid zonder duidelijke reden

GGADB202 : Vaak zenuwachtig

GGADB204 : Vaak hopeloos

GGADB207 : Vaak somber of depressief

GGADB210 : Vaak afkeurenswaardig, minderwaardig of waardeloos

Regie op het leven

GGRLB201 : Weinig controle over dingen die mij overkomen

GGRLB202 : Sommige van mijn problemen kan ik met geen mogelijkheid oplossen

GGRLB204 : Ik voel me vaak hulpeloos bij omgaan problemen van het leven

GGRLB206 : Wat er in de toekomst met me gebeurt hangt voor grootste deel van mezelf af

MCMZOS304 : Mantelzorg ontvangen

#### Inclusie-criteria

Uit de totale populatie (netto-respons) van de GGD-regio is aan de hand van onderstaande inclusiecriteria de groep (potentieel) kwetsbare inwoners geisoleerd:

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

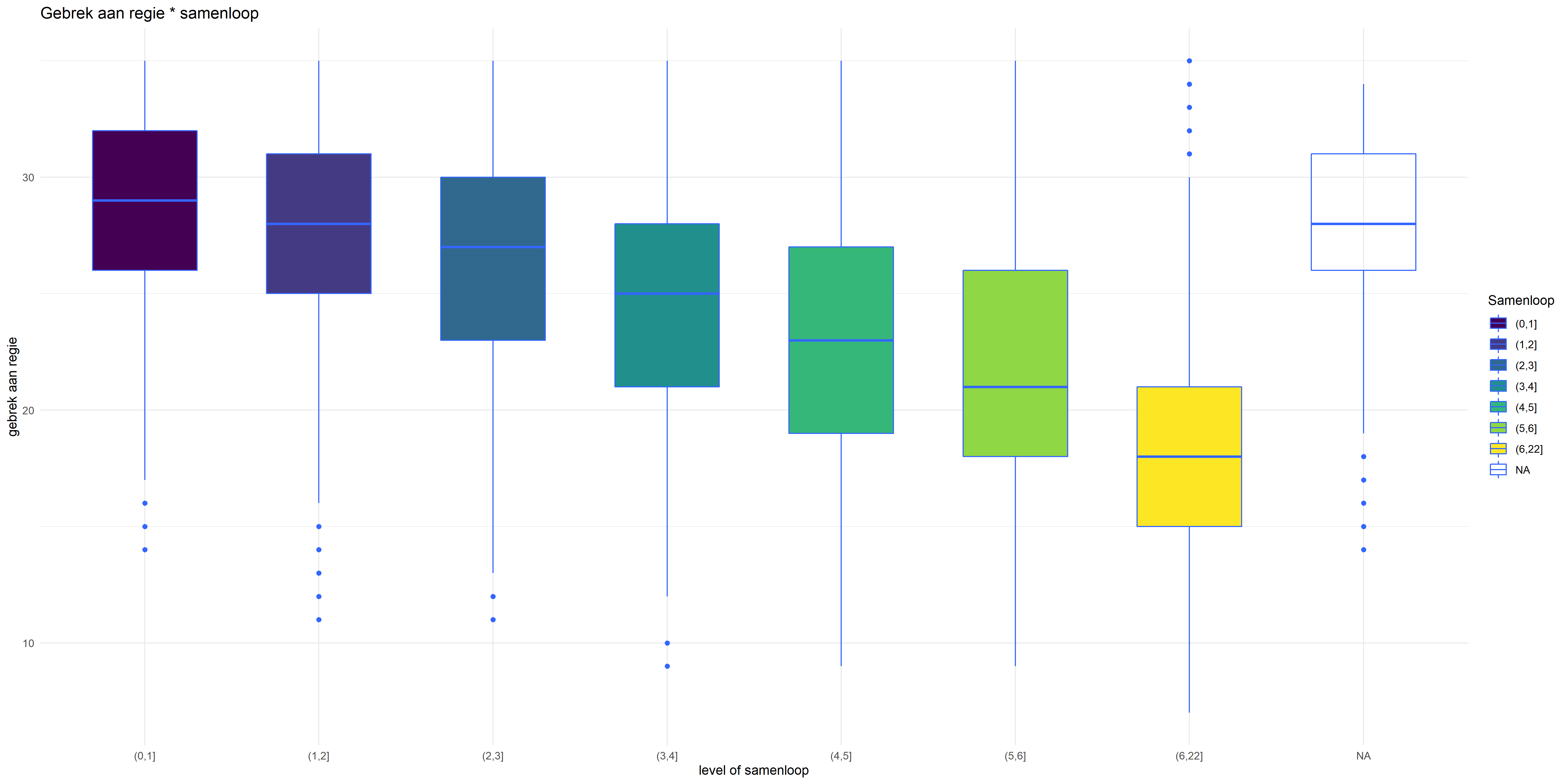
Gezondheidsbeleving: (zeer) slechte gezondheid (KLGGA207) : ==3

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

Daarmee kan 32% van de respondenten in meer of minder mate als kwetsbaar worden aangemerkt. Deze groep is in het vervolg van de analyse betrokken.

Een onder-/bovengrens voor respectievelijk de dimensies: samenloop van uitdagingen, gebrek aan regie, en angst & depressie, is afgeleid op basis van natural binning[[2]](#footnote-2):

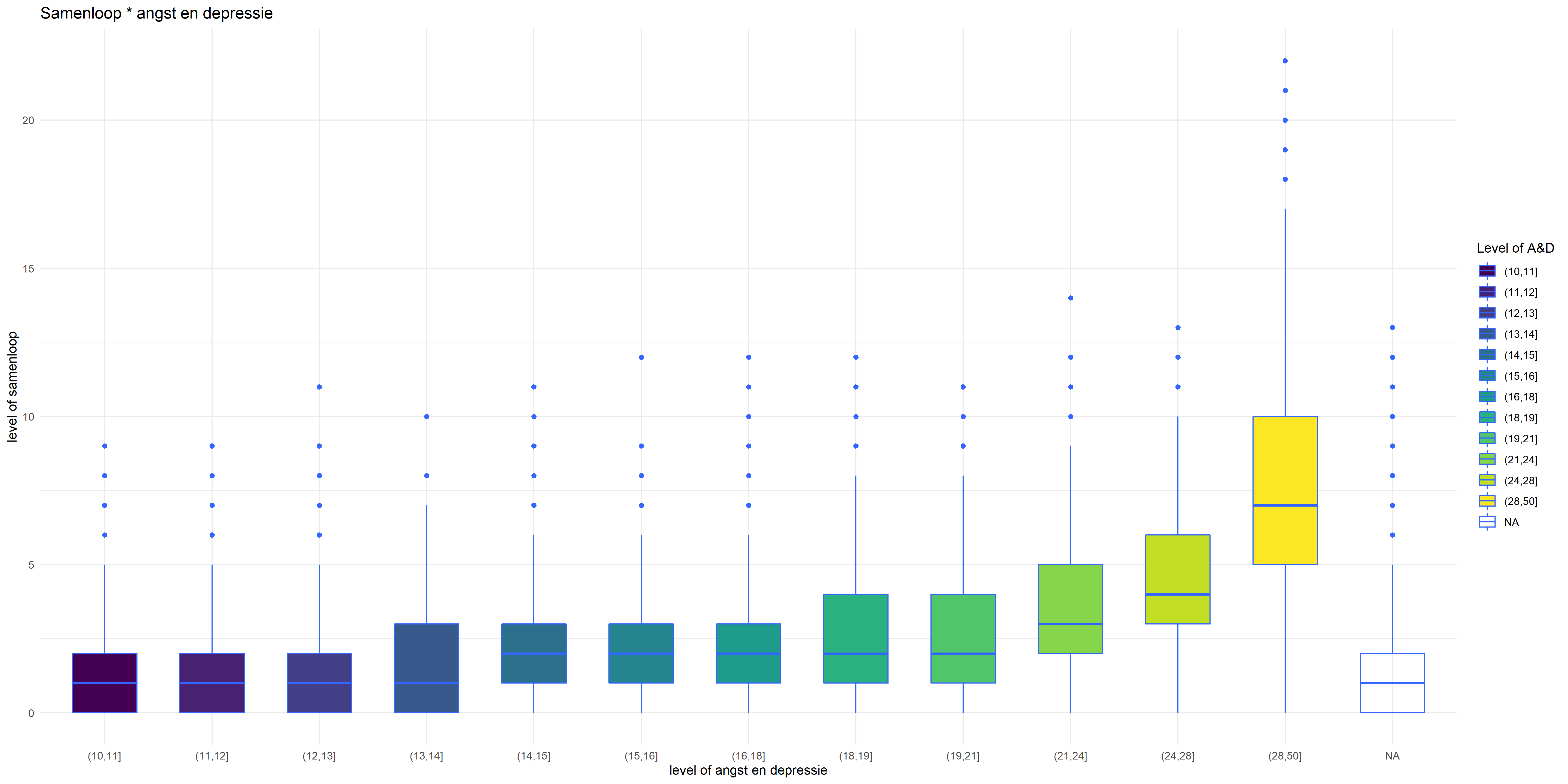
#### S*amenloop van uitdagingen*



#### Gebrek aan regie

#### 

#### Angst en depressie



### Imputatie van ontbrekende waarden

In dit onderzoek worden alle respondenten betrokken die aan de inclusive criteria voldoen. Ontbrekende waarden (missings) onder de featureset worden geschat zodat er sprake is van complete en accurate observaties per respondent. Daarvoor hanteren we een imputatietechniek die rekening houdt met het patroon in de featureset van gelijkwaardige observaties.

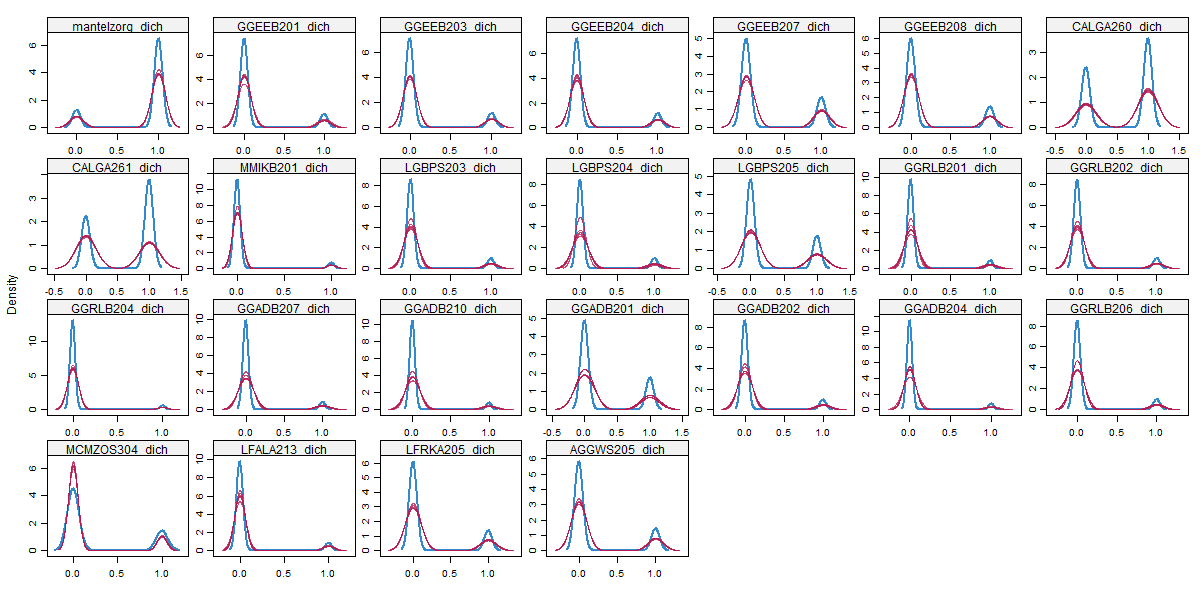
Methode : Multivariate Imputation via Chained Equations (MICE[[3]](#footnote-3)) aan de hand van logistische regressie voor binaire features.

Het MICE-algoritme bepaalt zelf welke features in hoge mate geschikt zijn voor het schatten van de ontbrekende waarden.

Een combinatie van 25 imputatiepogingen (iteraties) zijn nodig om te komen tot convergentie bij imputatie. Convergentie plots van (een deel van) de features:

Imputatie resultaat volgt in hoge mate de proporties onder de geobserveerde cases (resp. rode en blauwe lijn).

#### Geimputeerde waarden vs observaties



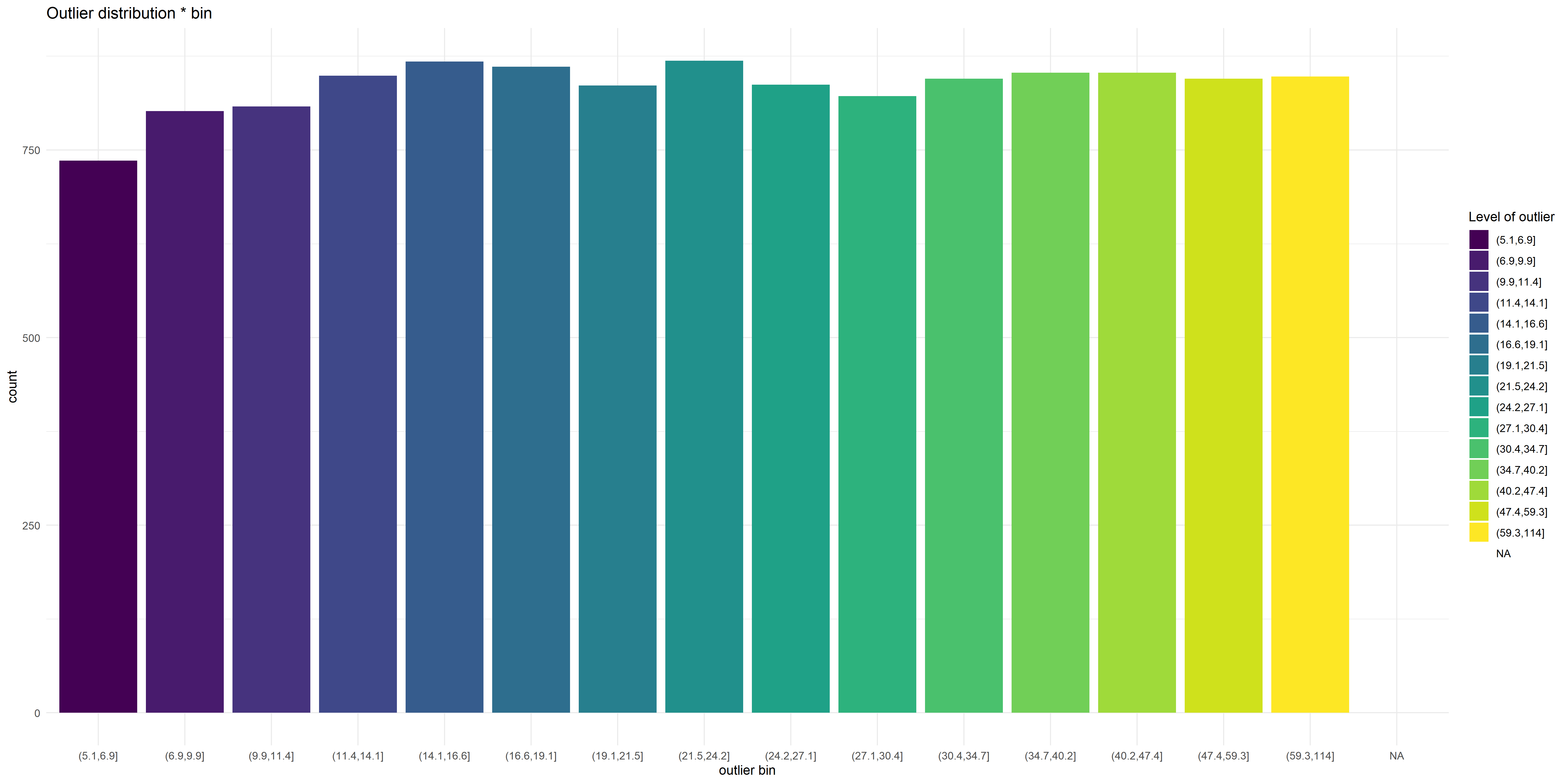
### Outliers

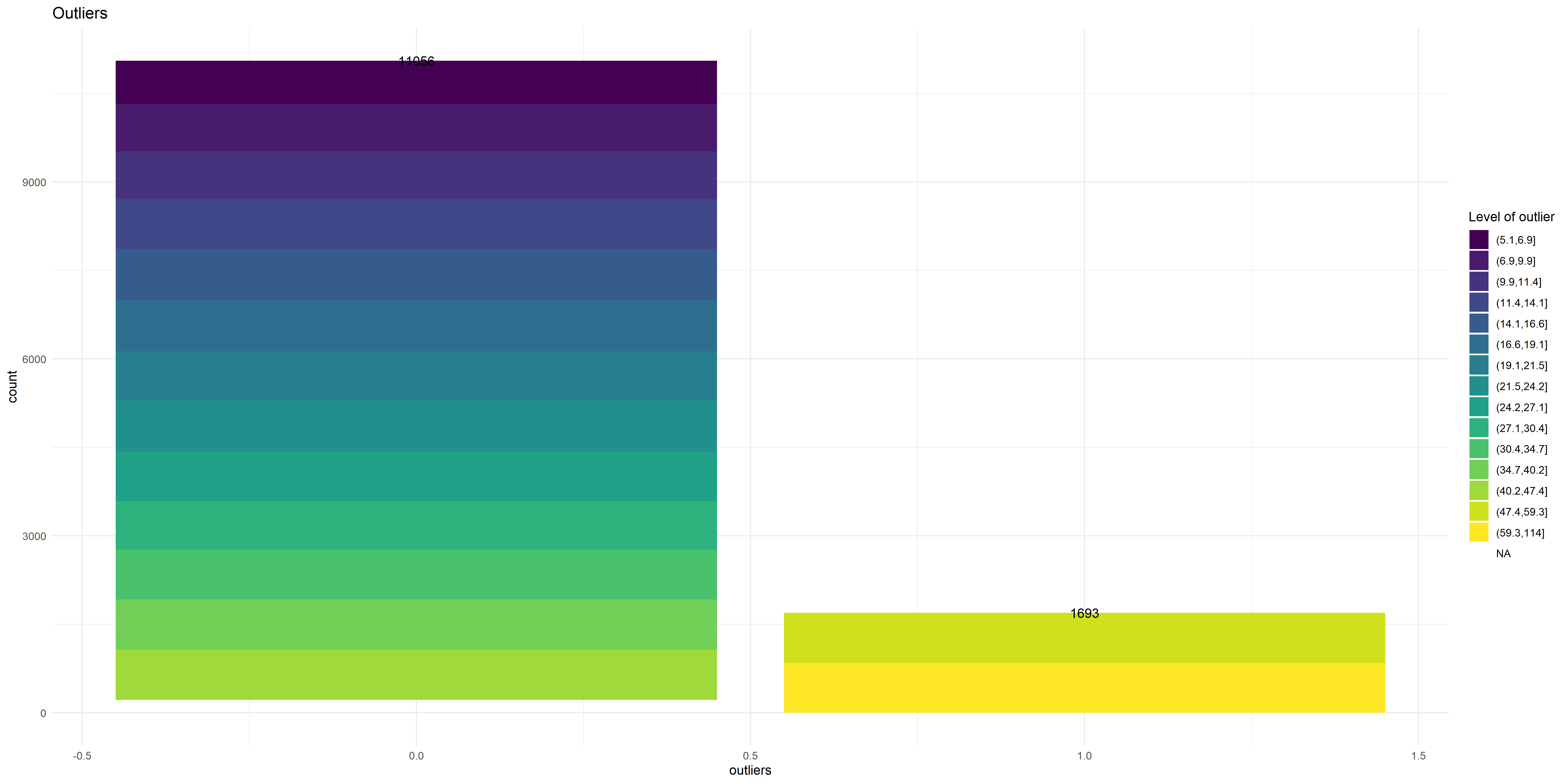
In het streven naar het herleiden van eenduidige verschijningsvormen (heldere patronen in de beleving onder kwetsbare inwoners) is het van belang om zicht te krijgen op de mate waarin extremen (outliers) zich voordoen.

Een outlier is hier een observaties die voor wat betreft het patroon in de featureset in hoge mate een uniek karakter kent.

Een geschikte methode voor outlier detectie voor hoog-dimensionale data is de Mahalanobis benadering.

Outliers zijn vervolgens op natuurlijke wijze gebinned in klassen. De gehanteerde outlier drempelwaarde (i.c. 47,4) is bepaald aan de hand van de ondergrens van de op een-na laatste bin.





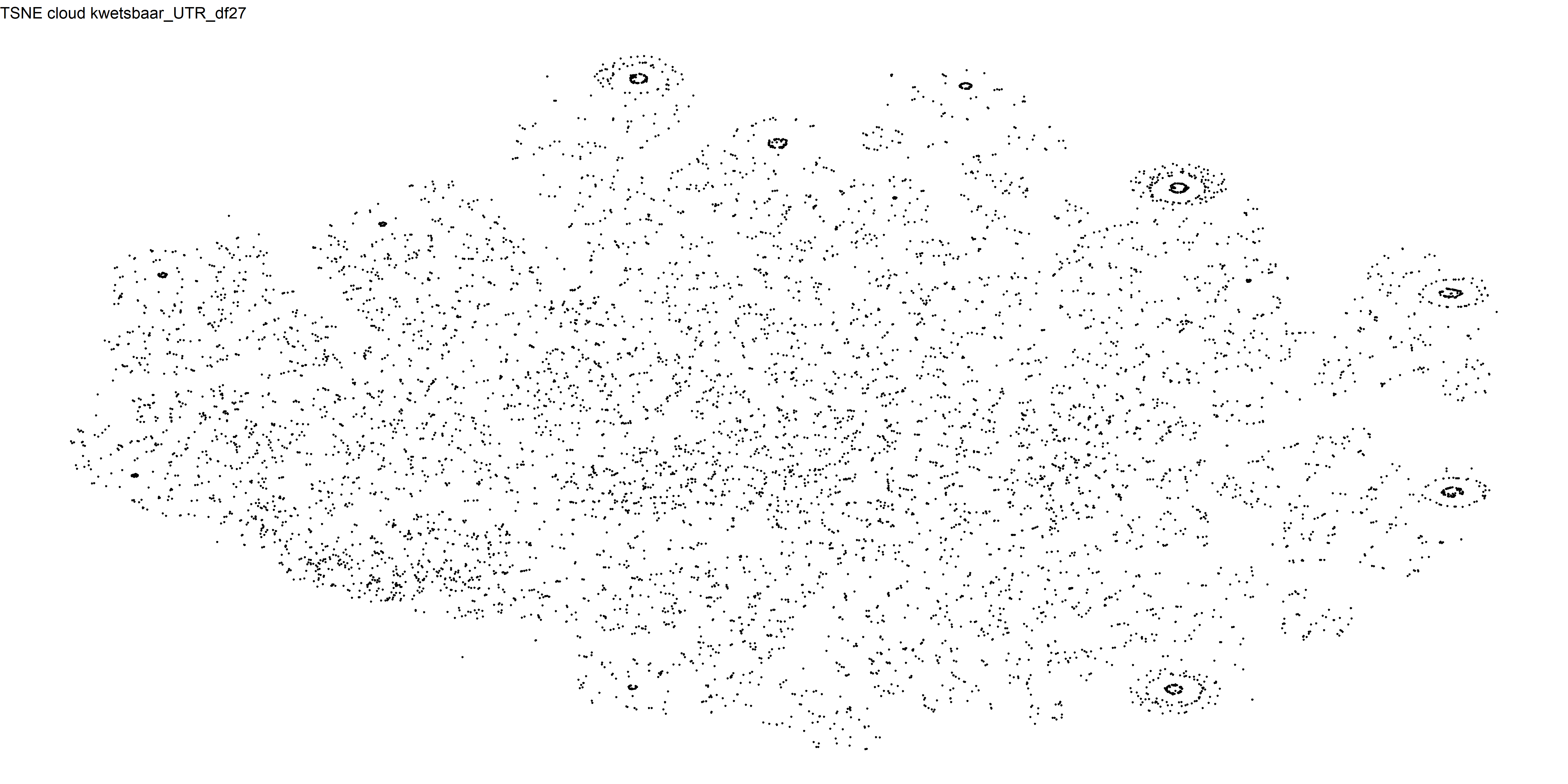
Van de totale populatie (37.495) zijn 1693 observaties aangemerkt als substantiele outliers.

### Dimensionaliteit reductie

Dimensionaliteit reductie is het proces om de dimensies van gegevens (i.c. samenstel van features) te verminderen waarbij zoveel mogelijk van de informatiewaarde blijft behouden. In dit geval vervatten we 27 features in 2 dimensies aan de hand van het TSNE algoritme[[4]](#footnote-4).

T-Distributed Stochastic Neighbor Embedding is een techniek voor dimensionaliteitsreductie die bijzonder geschikt is voor de visualisatie van hoogdimensionale gegevenssets.

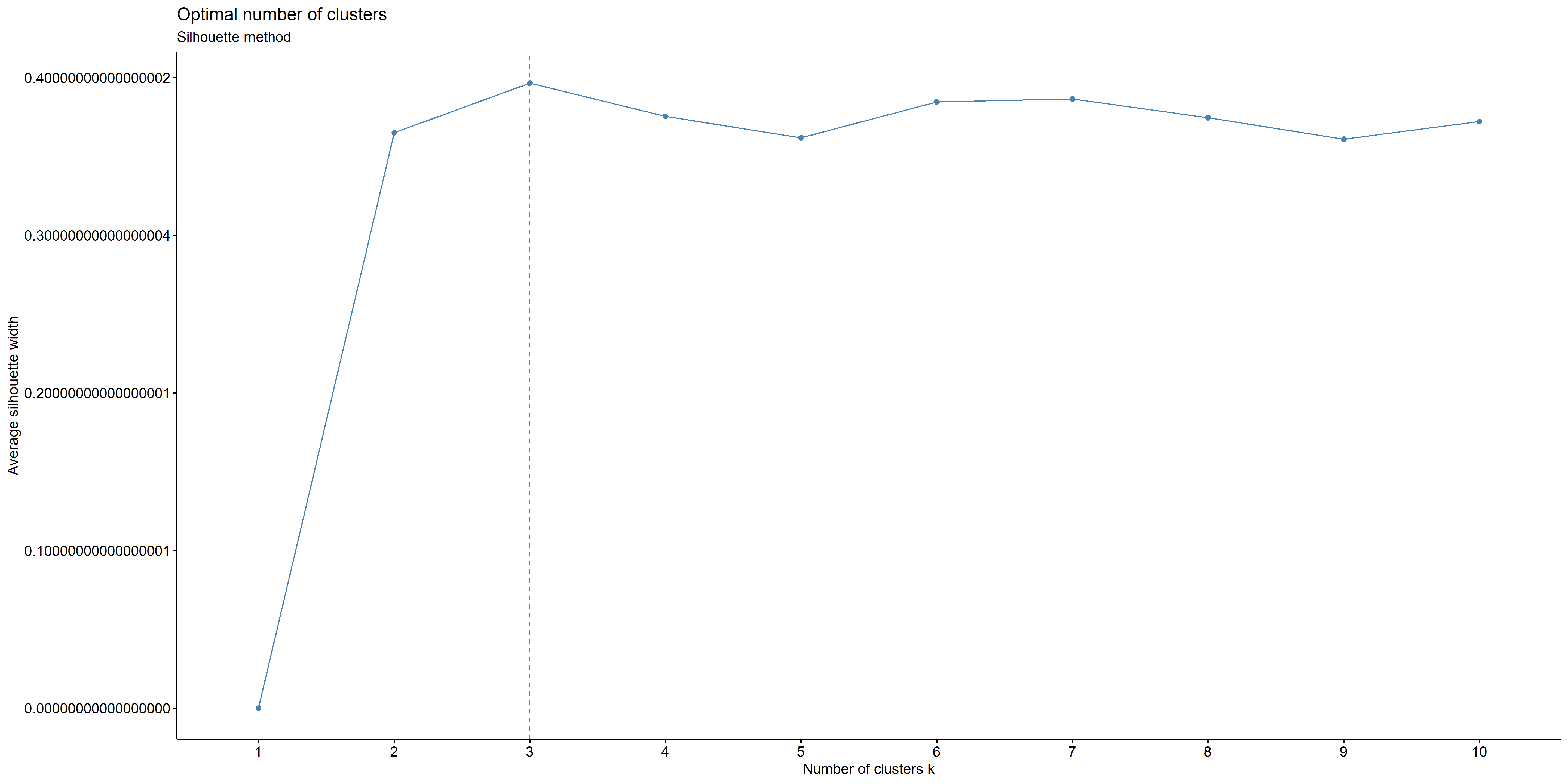
Het algoritme streeft naar het behoud van de afstand tussen punten uit de hoger dimensionale ruimte, waarmee punten die dicht bij elkaar liggen verwijzen naar vergelijkbare beleving.



De visualisatie kan worden gezien als een radar waarop de kwetsbare personen ten opzichte van elkaar zijn afgetekend.

### Clustering

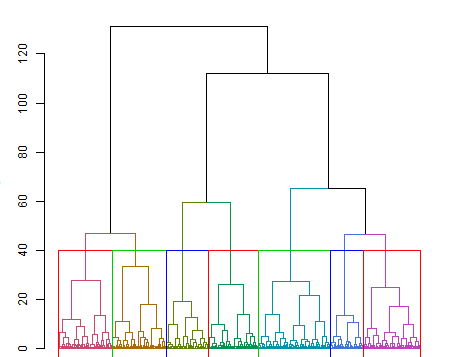
Aan de hand van een GAP/Silhoutte-analyse is het aantal clusters bepaald. In dit onderzoek is het doel niet zozeer om een optimal aantal clusters tot stand te brengen (zoveel mogelijk informatiewaarde en differentiatie in zo min mogelijk clusters), maar een handzaam aantal verschijningsvormen (tussen 5 en 10 clusters) met een relatief hoog onderscheidend vermogen.



Het aantal clusters is vastgesteld op 7 als eerstvolgende -gelijkwaardig- alternatief voor het optimale aantal clusters (3). Het vervolg van deze analyse is gebaseerd op 7 clusters.

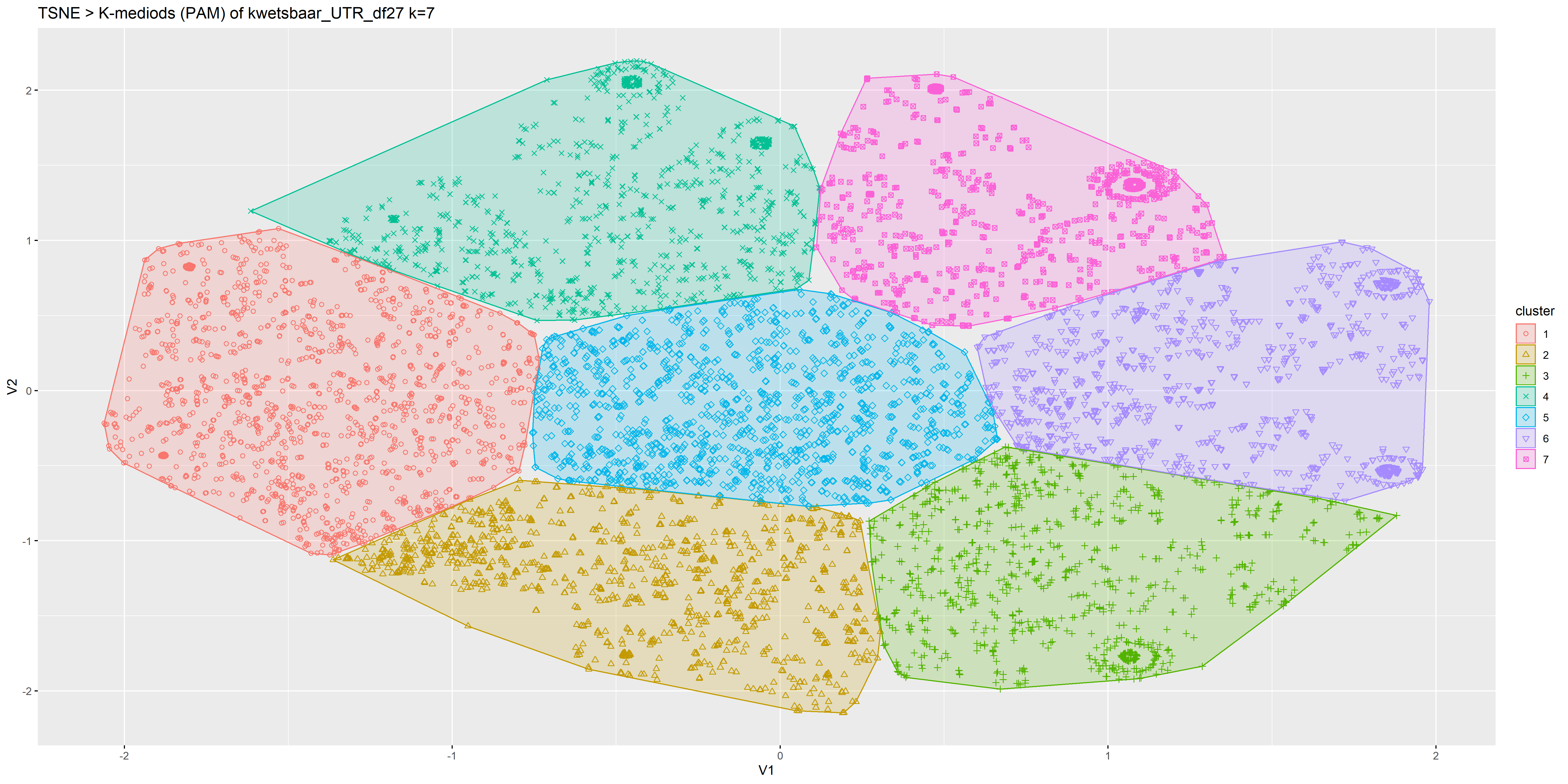
#### Hierarchische clusteranalyse

De eerste clustermethode betreft de hierarchische clusteranalyse om te verkennen hoe de clusters zich ontvouwen met een steeds groter aantal clusters (2:k). Dit moet enig geval geven bij de keuze voor 7 clusters. De zeven clusters in kwestie zijn met kleur gemarkeerd.



#### K-mediods clustering (PAM)

Vervolgens is aan de hand van het K-mediods clusteralgoritme (PAM) gekeken naar een meer algemene wijze van clusteren. K-mediods streeft er naar om de ongelijkheid tussen punten in een cluster te beperken. K-medoids gebruikt een feitelijke observatie om een cluster te construeren. Een medoid 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. In het geval van GGD regio Utrecht hebben we te maken met een substantiele groep outliers.

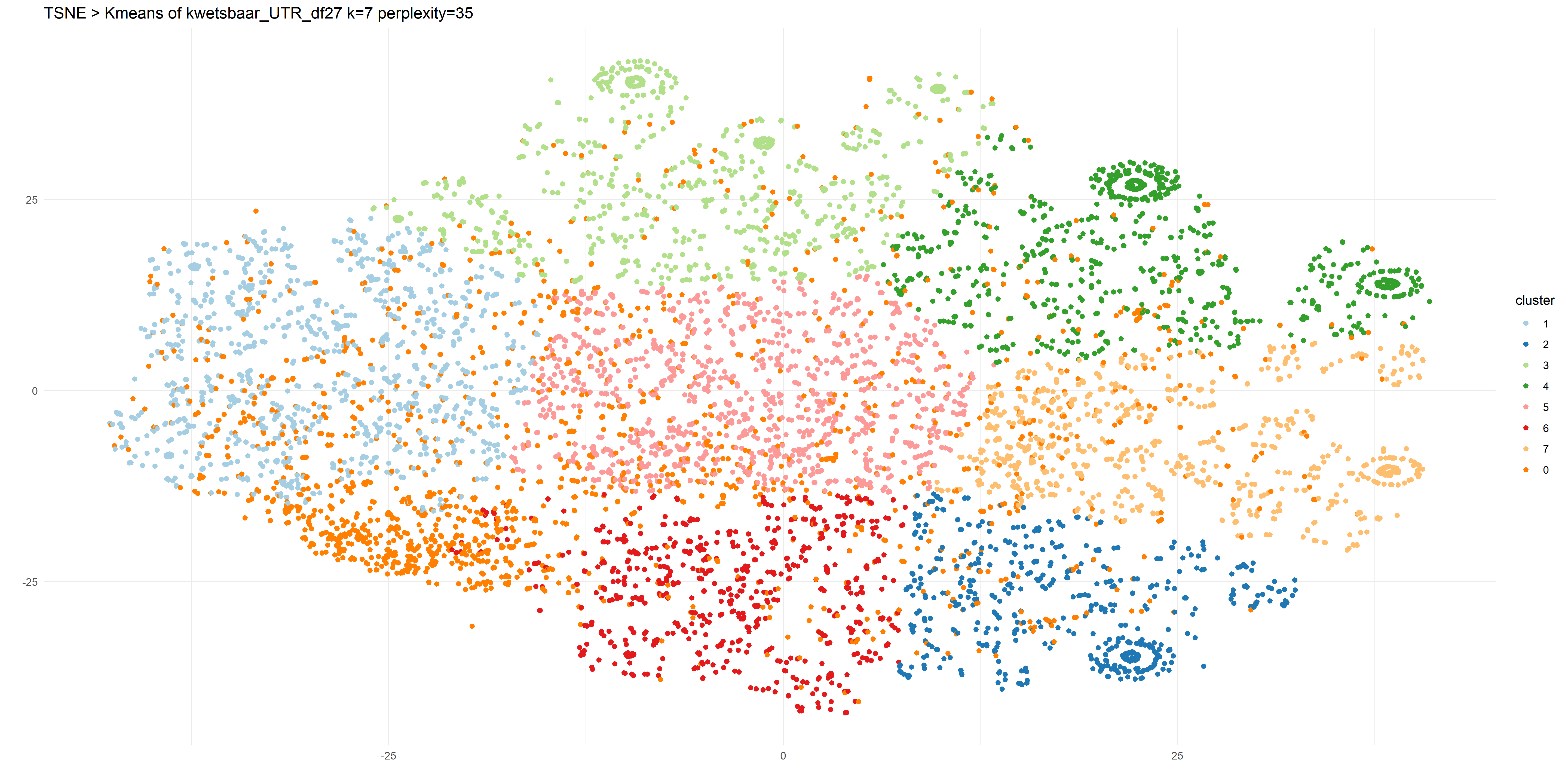


De clusters geven de overeenkomstige *ervaren situatie van kwetsbaarheid* weer. De observaties (punten) zijn de individuele kwetsbare inwoners.

#### K-means

In tegenstelling tot PAM streeft K-means naar het beperken van de afstand van een observatie tot het centrale punt in een cluster (niet-noodzakelijkerwijs een daadwerkelijke observatie). 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 indeling te verkrijgen worden in deze K-means analyse de outliers buiten beschouwing gelaten (apart genomen in groep ‘0’).

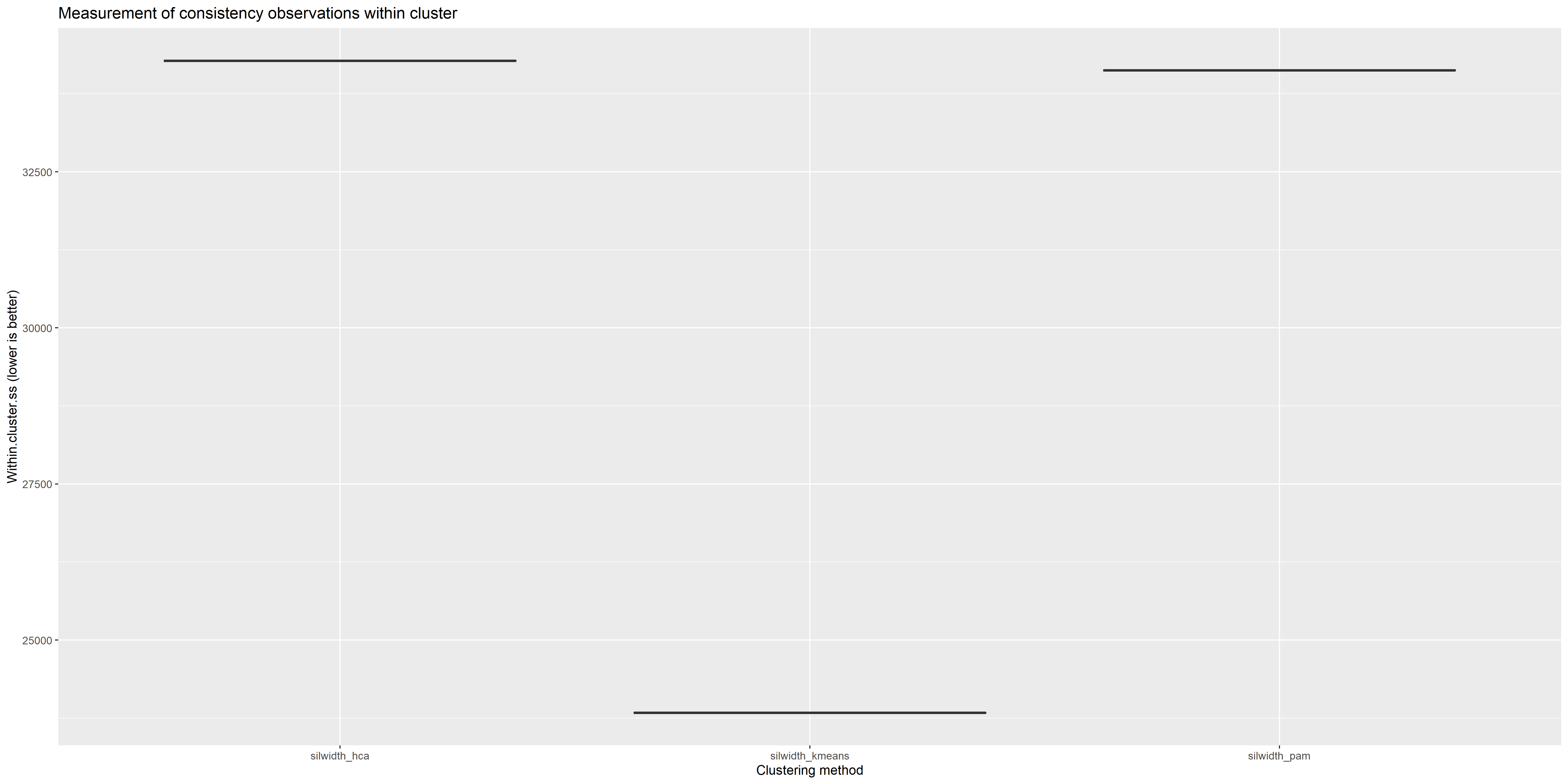


De observaties nabij een naburig cluster vertonen in hogere mate overeenkomsten met het betreffende buurcluster. De observaties uit groep ‘0’ (extremen / outliers) liggen verspreid over meerdere clusters, maar kennen wel een zeker mate van concentratie. Het betreft geen uniform cluster (of uberhaupt een cluster in de zienswijze van de gehanteerde clustering algoritmen).

Let op: De nummering van de clusters worden voor resp. de K-means en K-mediods eigenstandig bepaald. B.v. Cluster 1 van K-means is qua lading niet (soort)gelijk aan cluster 1 van K-mediods.

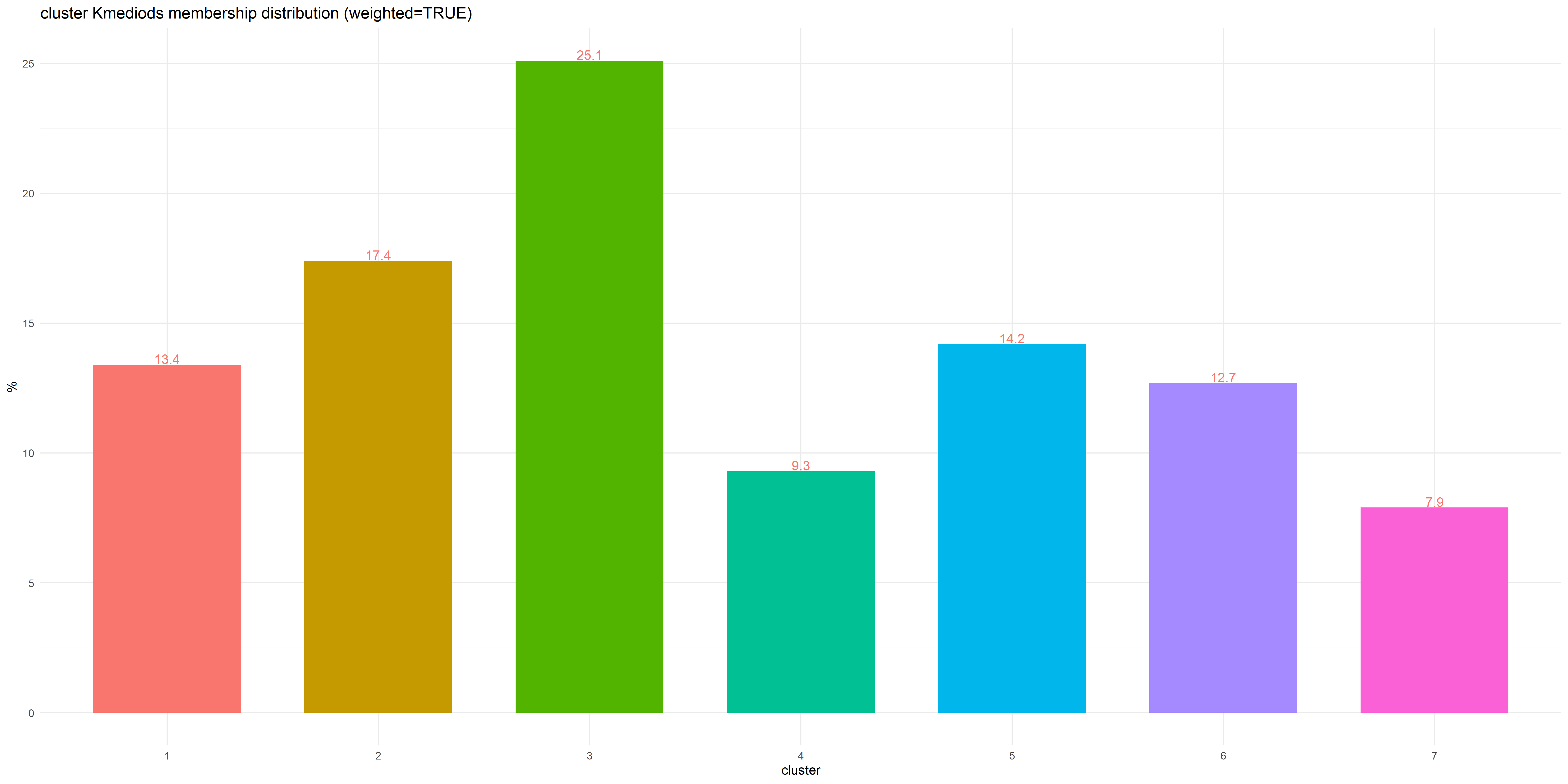
### Selectie methode

Vergelijking van de drie gehanteerde clustermethoden: hierarchische clusteranalyse (HCA), K-Mediods (PAM), en K-means. Lagere waarde duidt op een hogere samenhang binnen clusters. K-means scoort het beste, gevolgd door resp. K-mediods (PAM) en HCA. De reden dat K-means beter scoort ligt in het feit dat de ‘echte’ outliers buiten het process van clustering zijn gelaten. Dat is een expliciete keuze geweest om een stabiele Kmeans analyse tot stand te brengen en meer zich te krijgen op de ligging van ‘echte’ outliers.



De resultaten zijn vervolgens gesegmenteerd op basis van de K-mediods methode omdat deze populatie een grote groep outliers kent.

#### Cluster lidmaatschap distributie[[5]](#footnote-5)



#### Cluster lidmaatschap distributie \* gemeente

#### 

#### Clusterbeschrijving

In eerste aanleg zijn de clusters beschreven naar resp. leeftijd en de outcome dimensies.

#### Gemiddelde scores (gewogen)

cl\_pam leeftijd\_w samenloop\_w eenzaam\_w regie\_w depri\_w

*<int>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>*

1 65.1 6.62 4.74 19.8 25.2

2 44.5 5.82 5.44 21.5 26.7

3 38.1 2.35 5.31 24.0 21.5

4 64.3 3.88 4.58 21.5 20.6

5 56.0 5.84 7.22 22.3 22.8

6 50.8 2.11 5.66 23.2 19.3

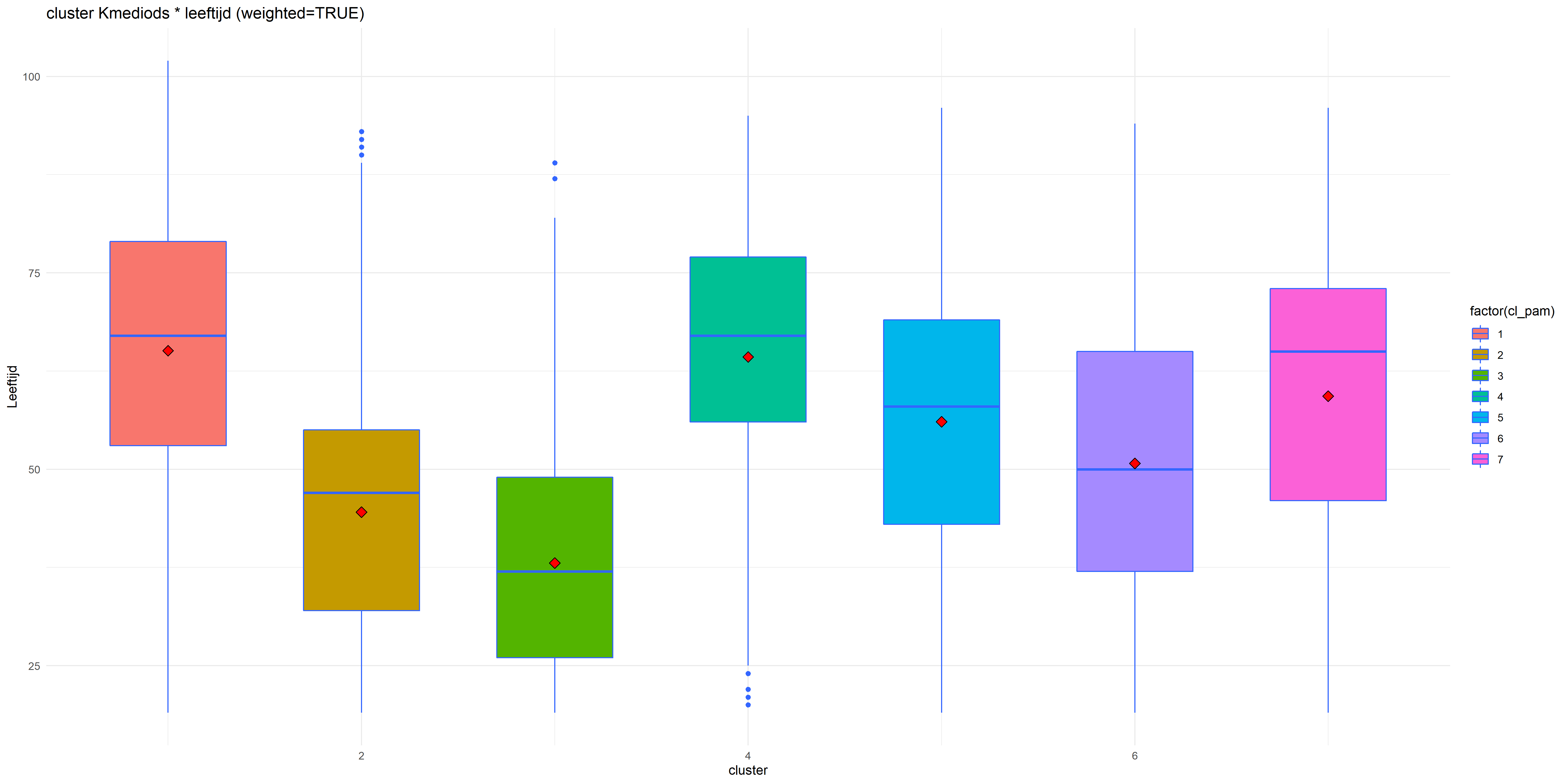
7 59.3 1.72 4.96 22.1 19.6

NA 46.5 1.23 1.53 29.6 13.6

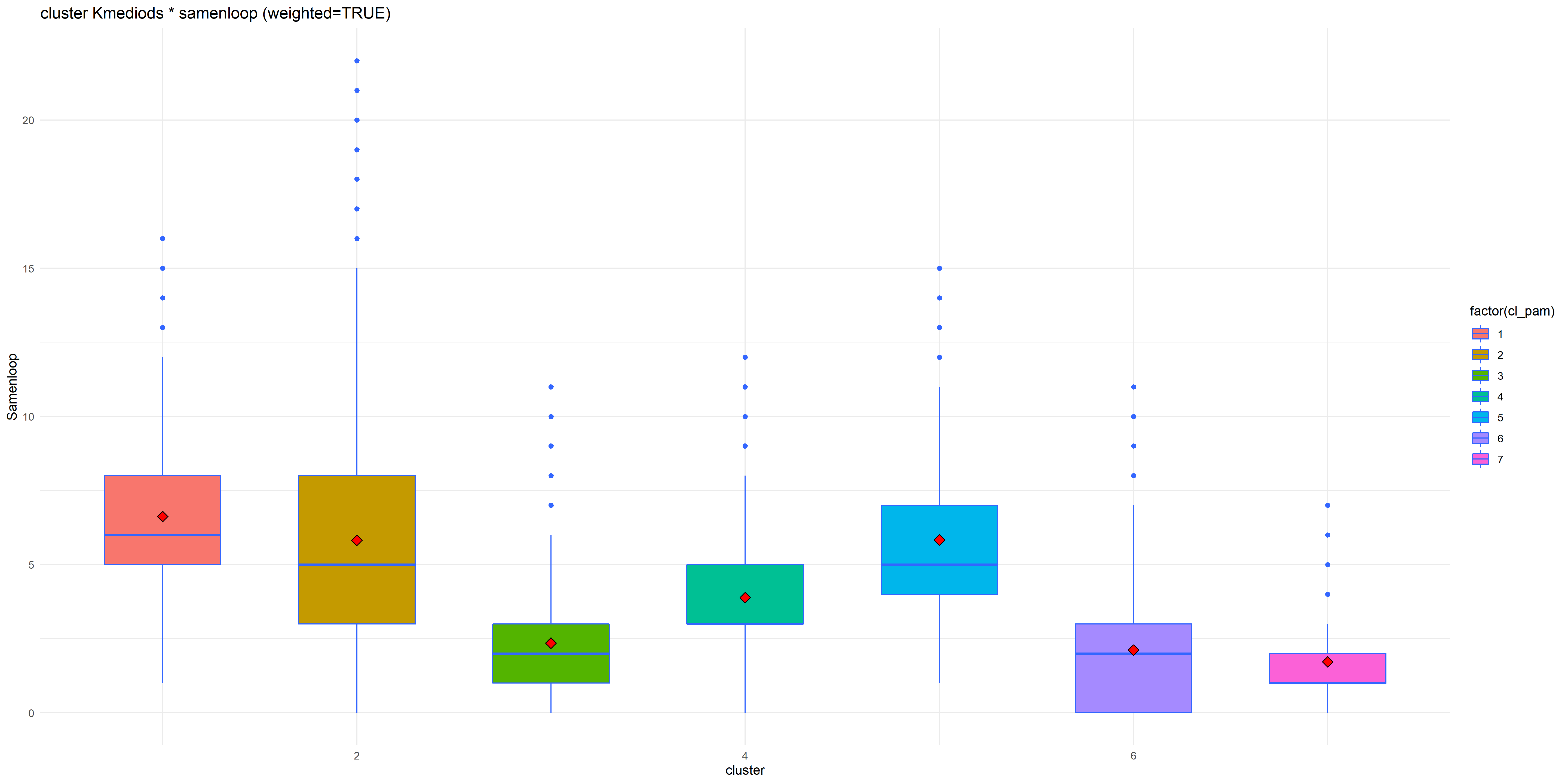
#### De groep ‘NA’ vertegenwoordigd de ‘niet-kwetsbare’ bevolking

In onderstaande boxplots vertegenwoordigt de rode punt het *gewogen* gemiddelde, de streep de mediaan, de blauwe punten de outliers.

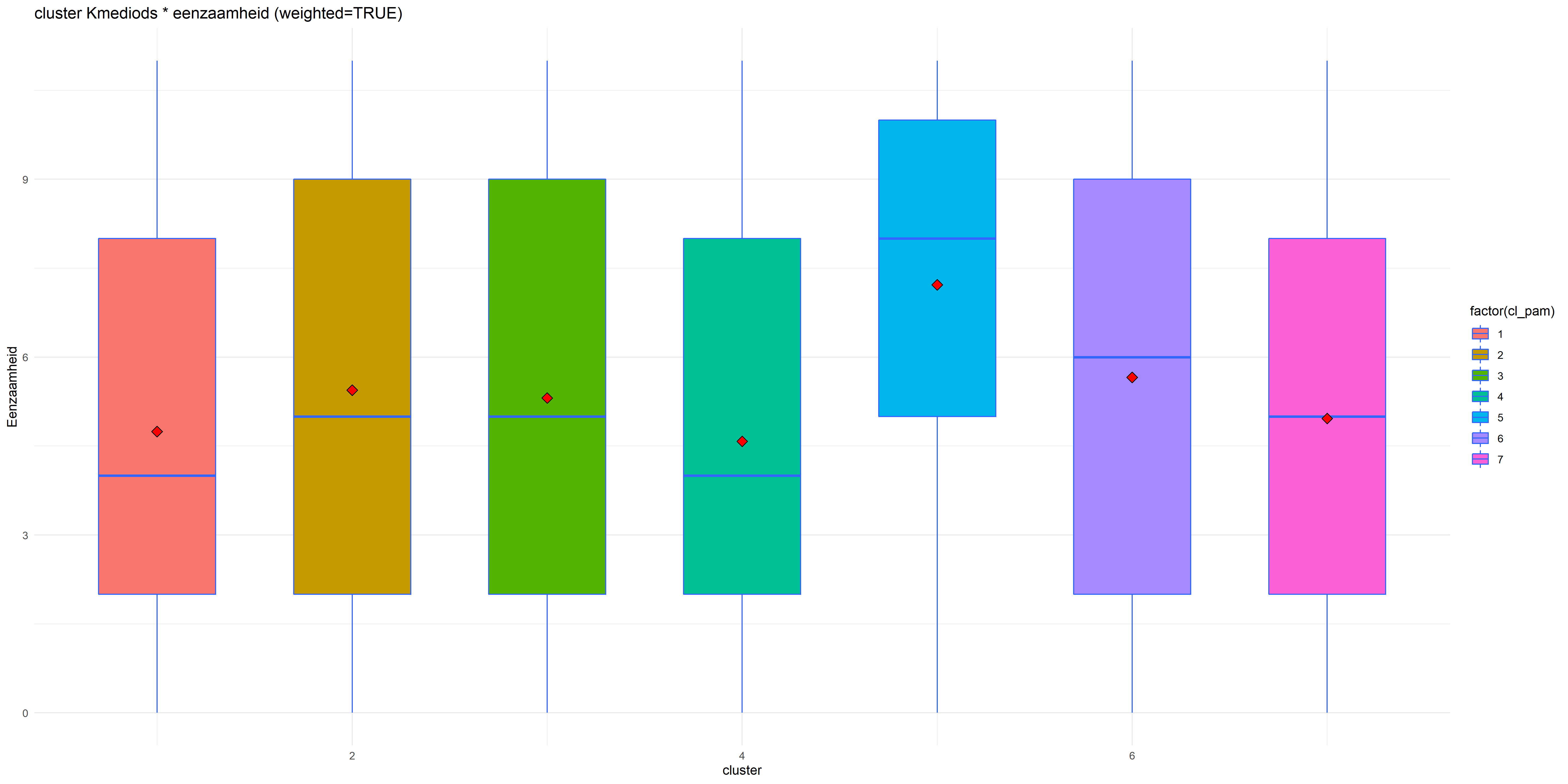
#### Leeftijd



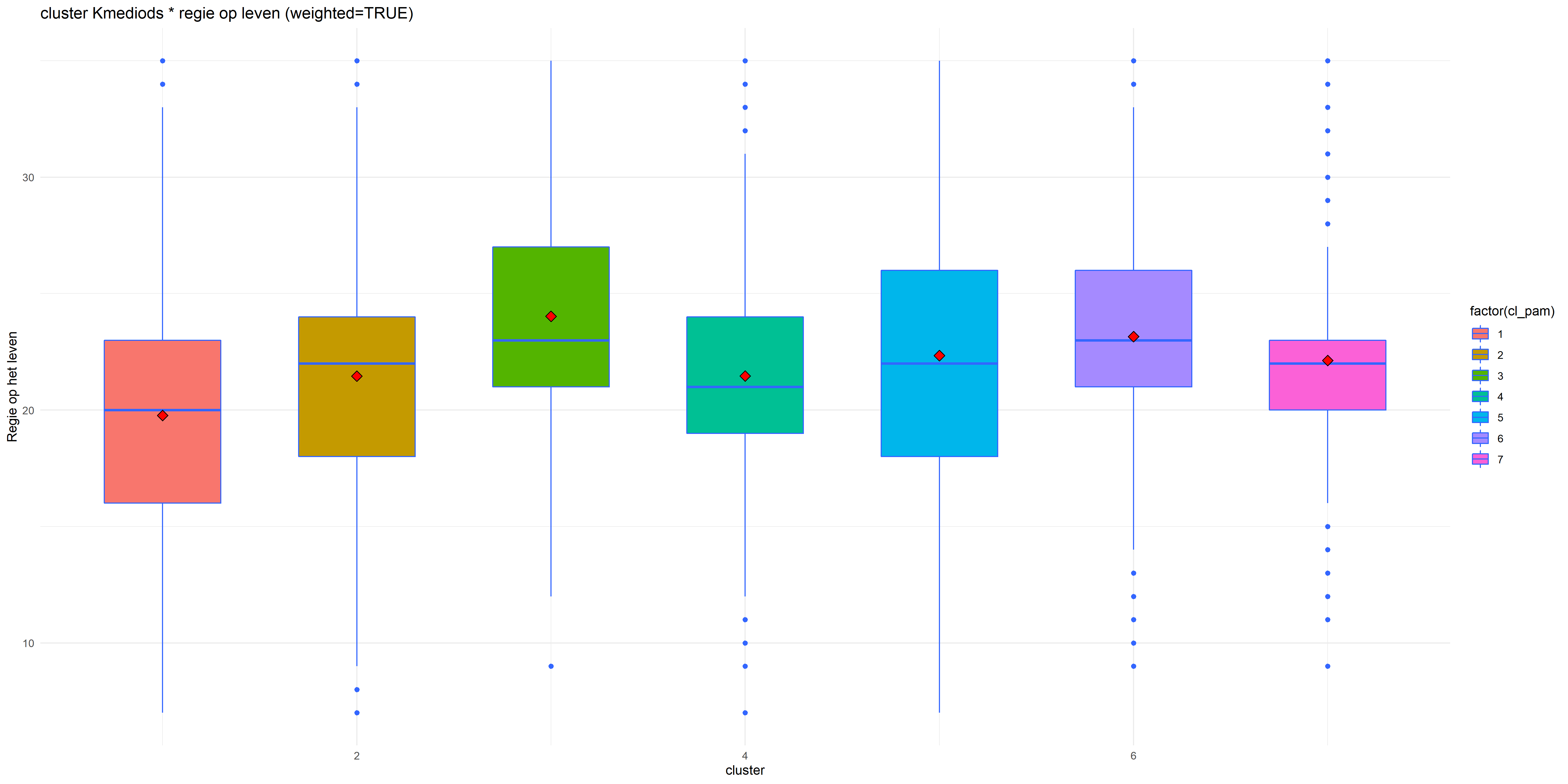
#### Samenloop van uitdagingen



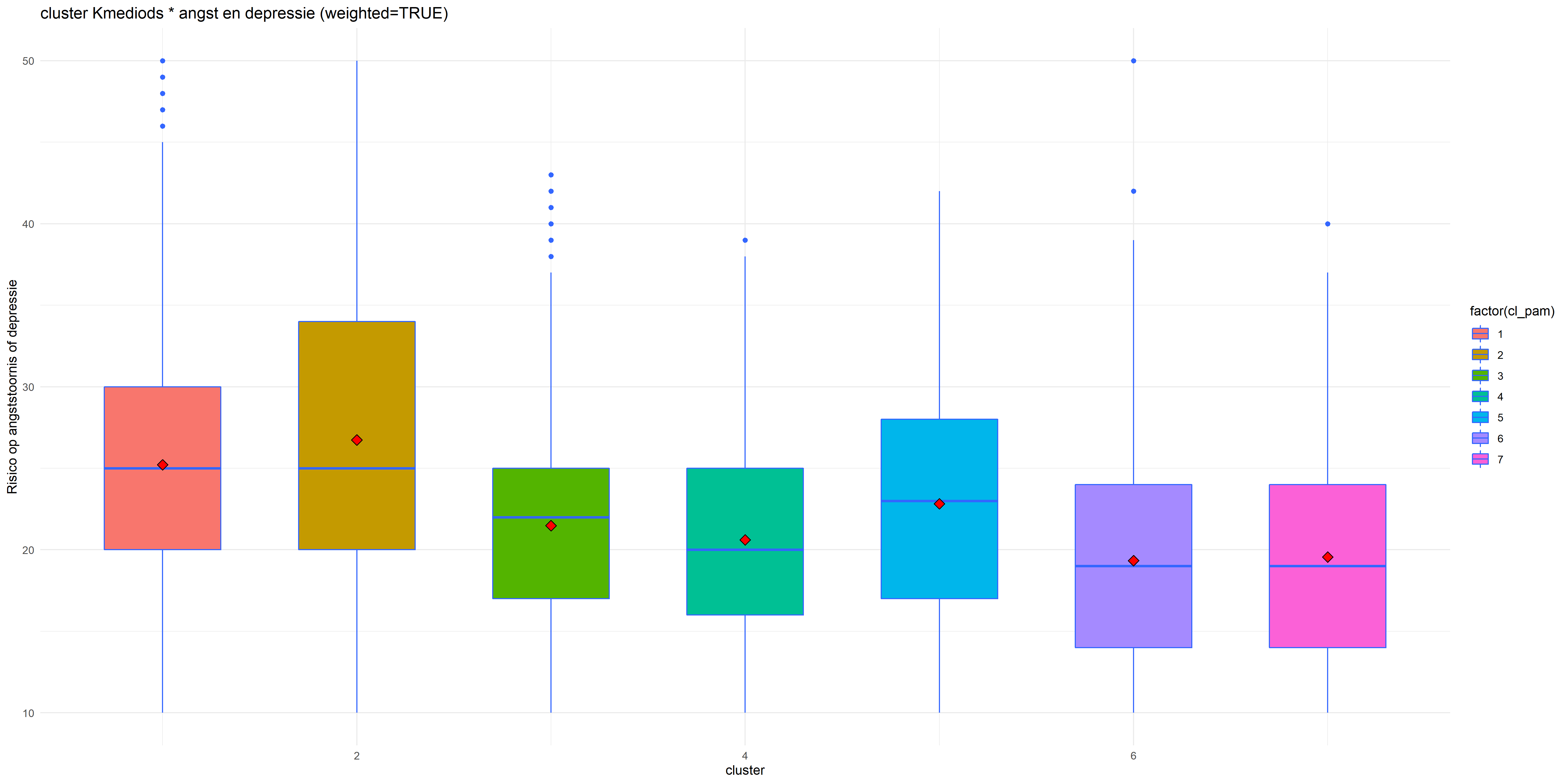
#### Eenzaamheid



#### Regie op het leven

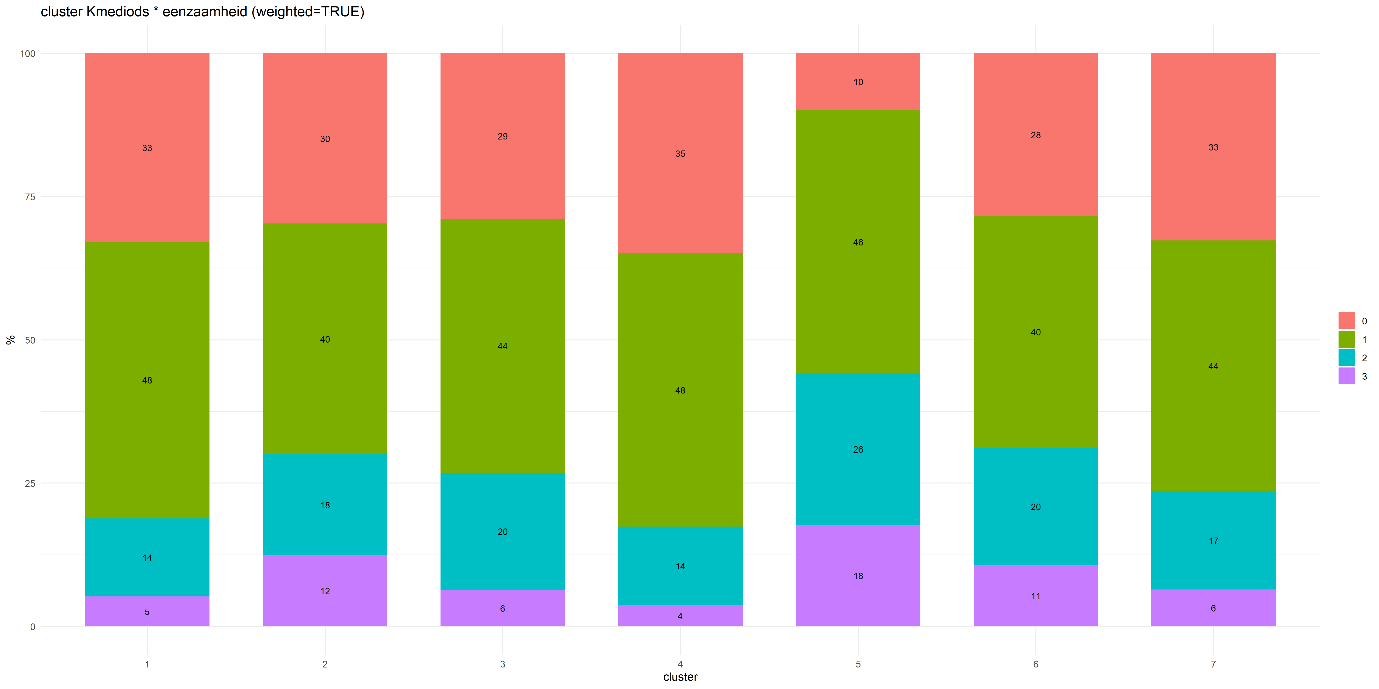


#### Angst en depressie

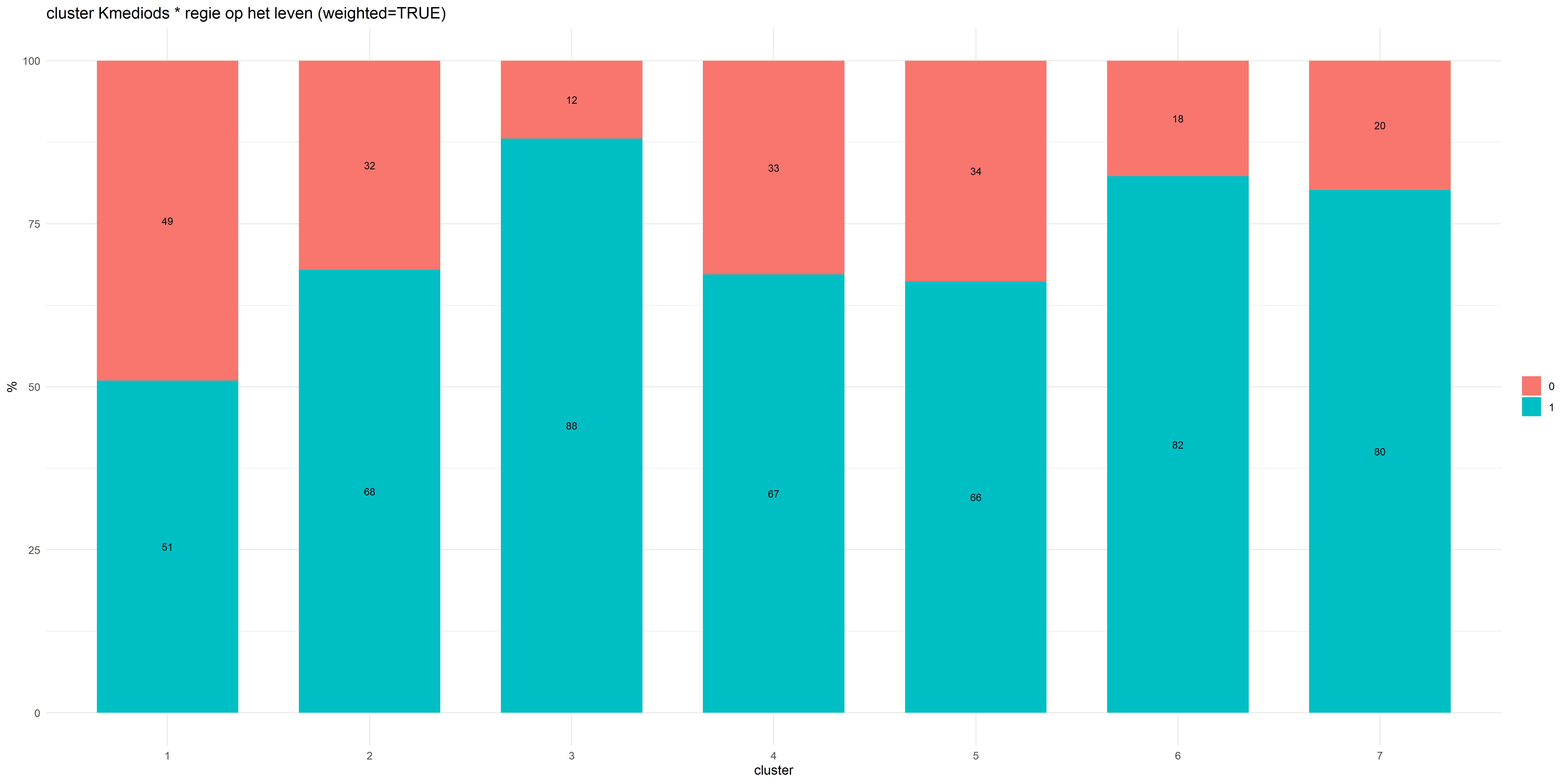


### Secundaire segmentatie

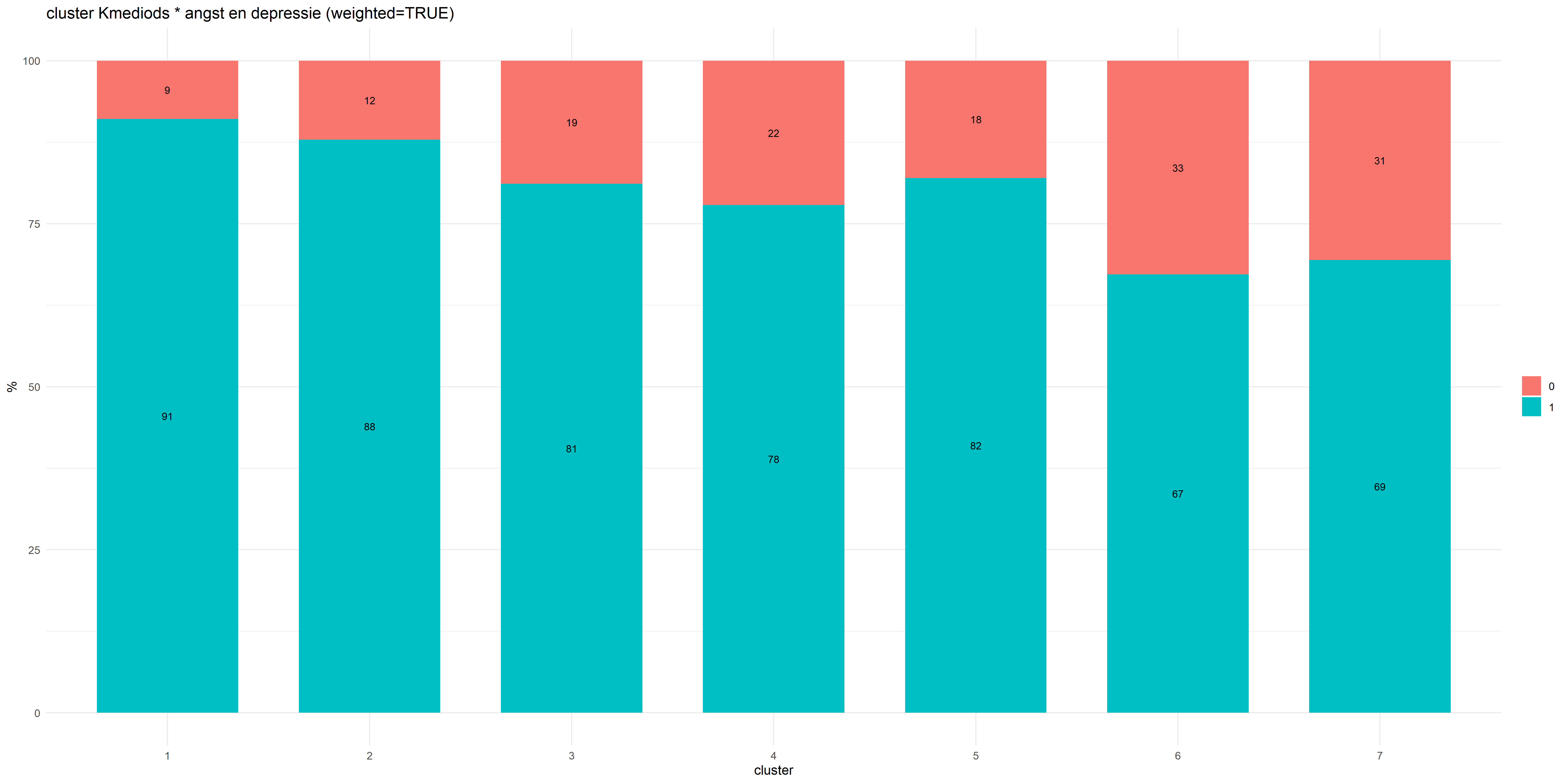
#### Eenzaamheid (GGEES208)



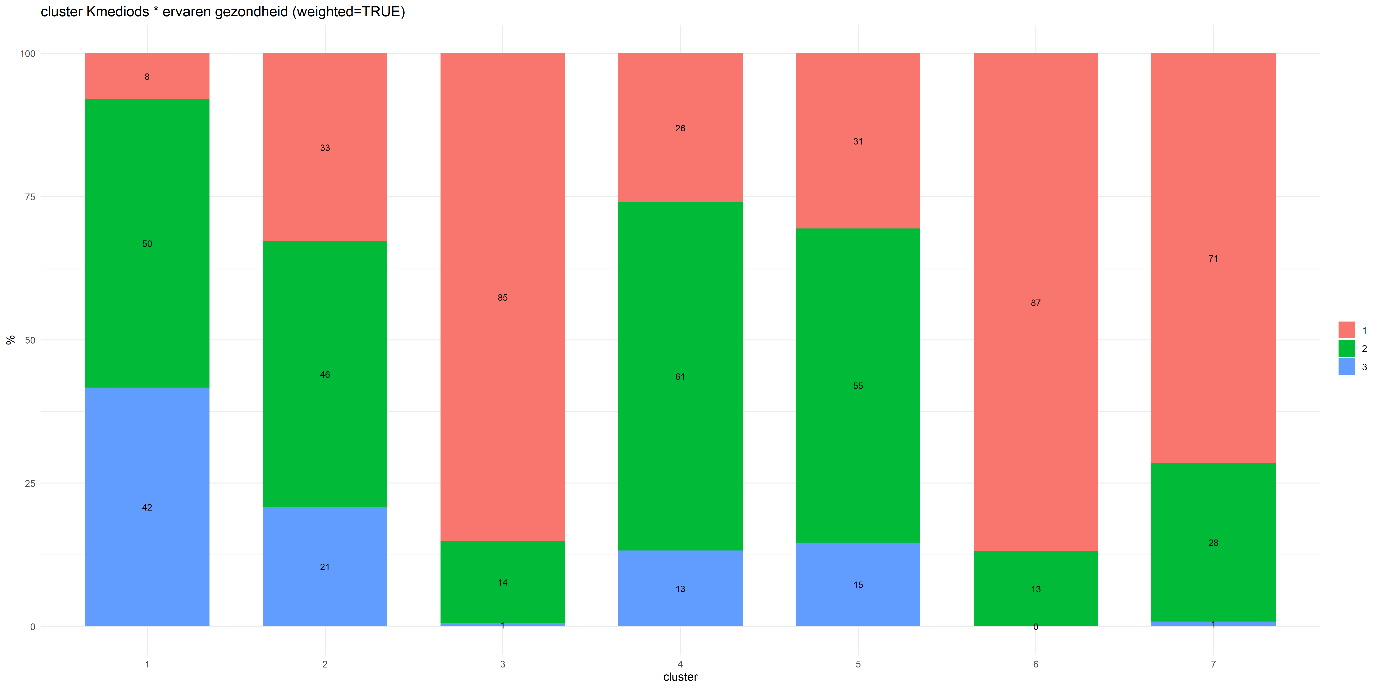
#### Regie op het leven (matig tot veel) (GGRLS203)



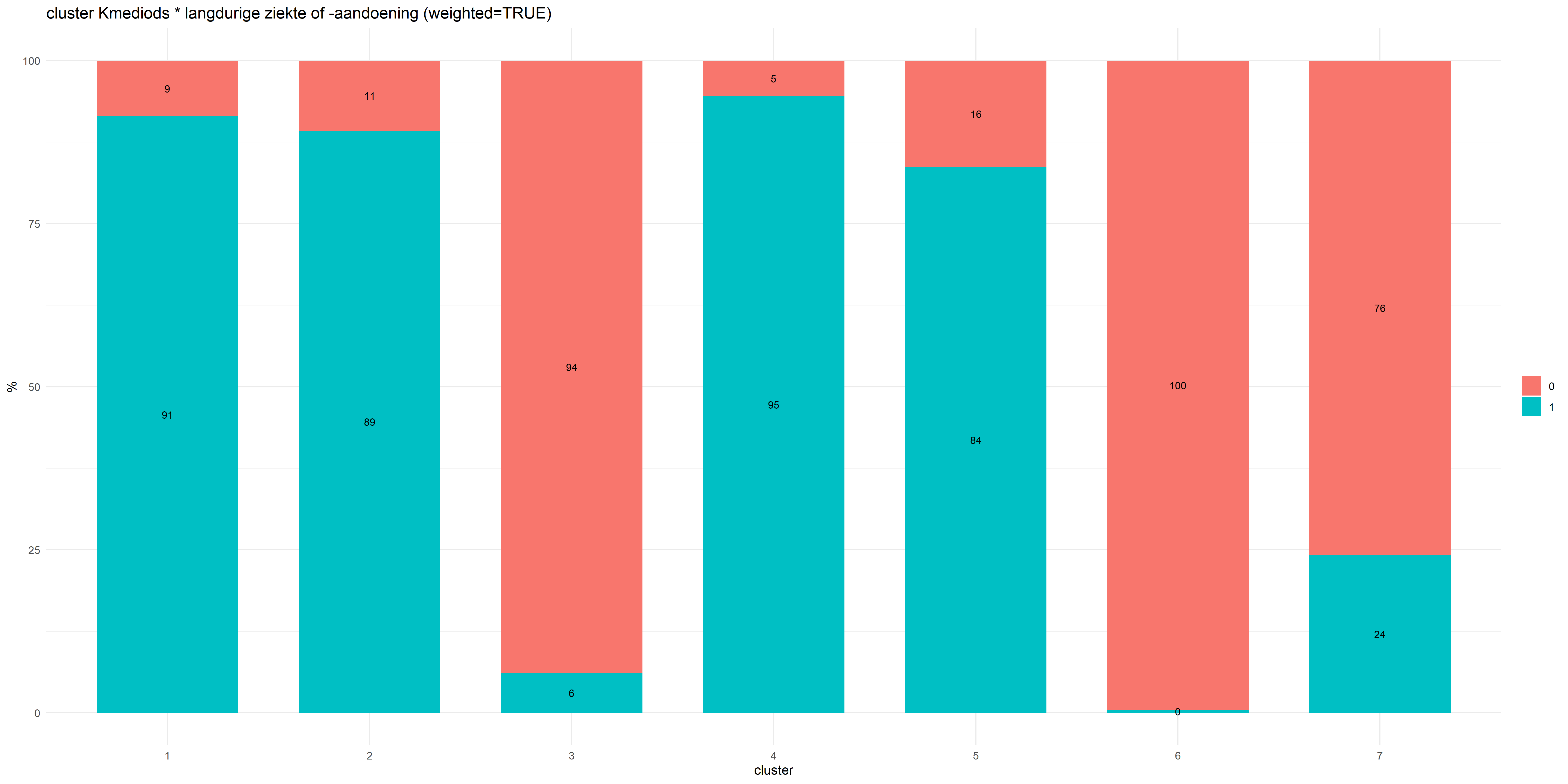
#### Angst en depressie (matig tot hoog risico) (GGADA202)



#### Ervaren gezondheid (KLGGA207)



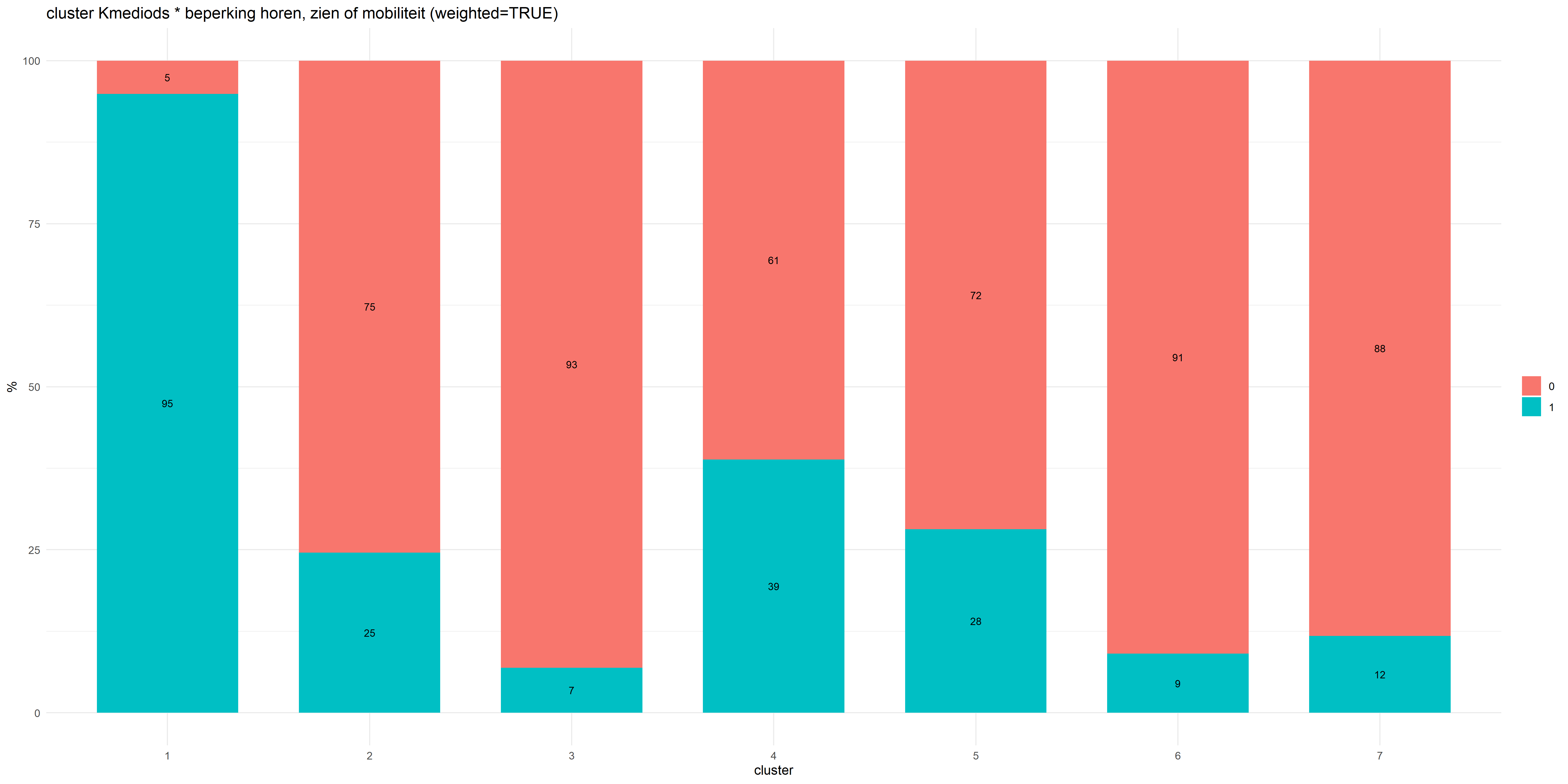
#### Langdurige ziekte of - aandoening (CALGA260)



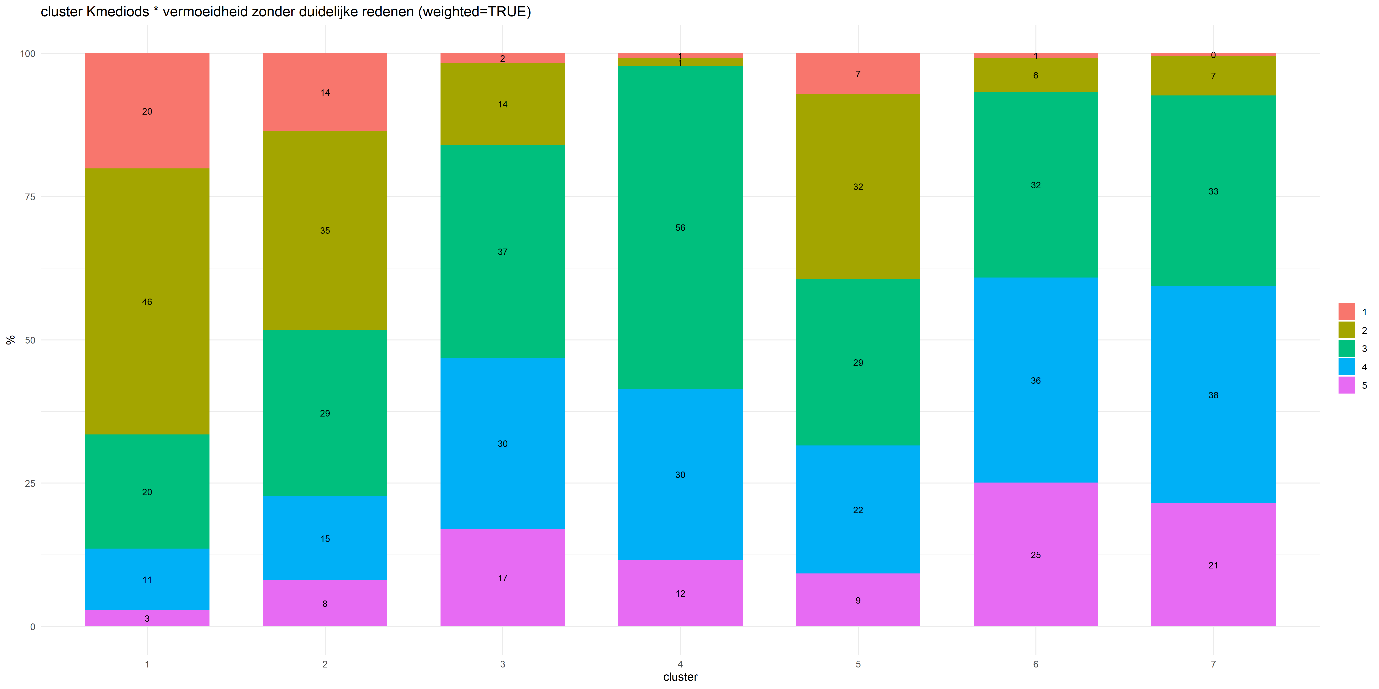
#### Beperking gezondheid (CALGA261)

#### 

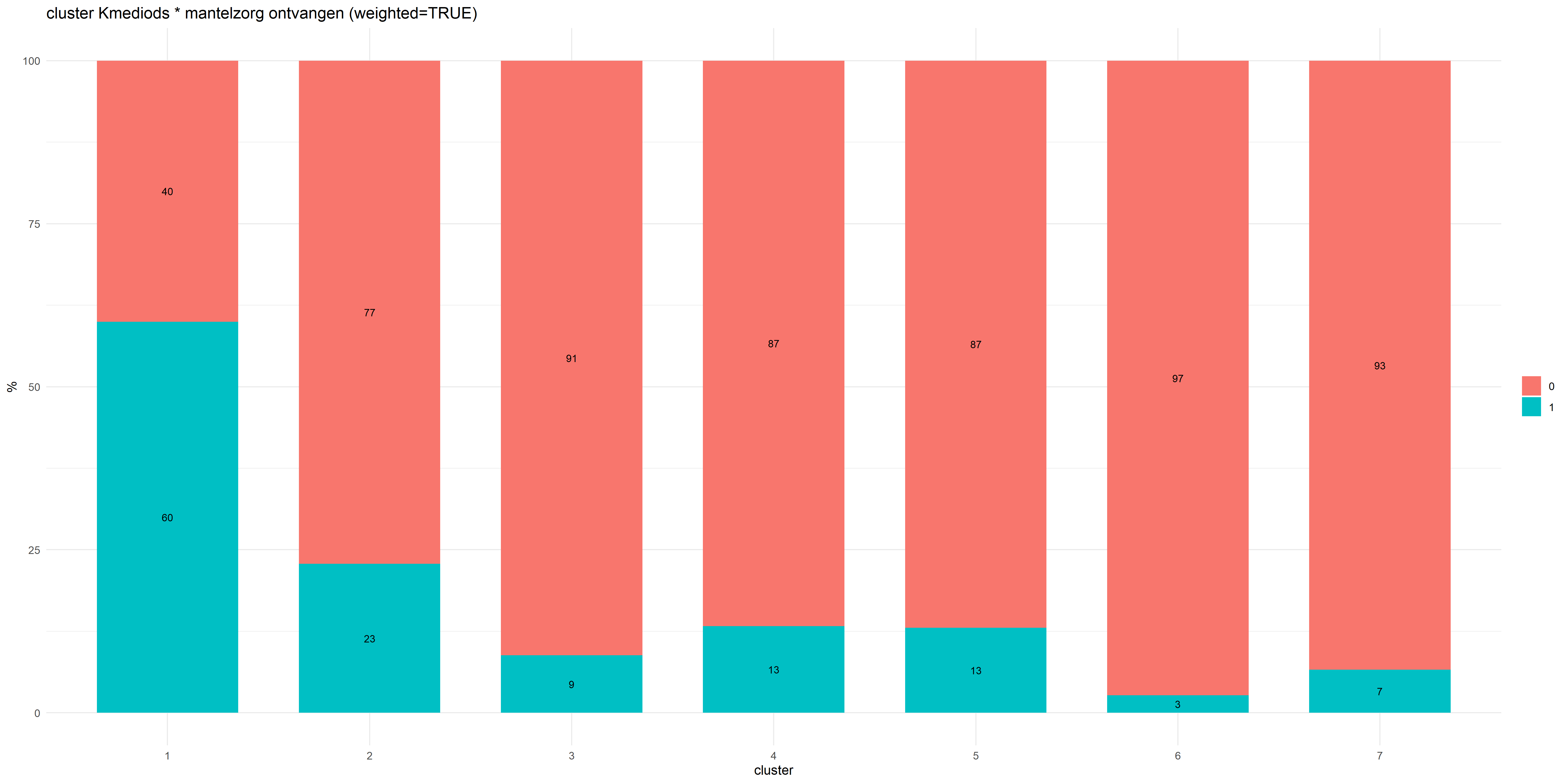
#### Beperking horen, zien of mobiliteit (LGBPS209)



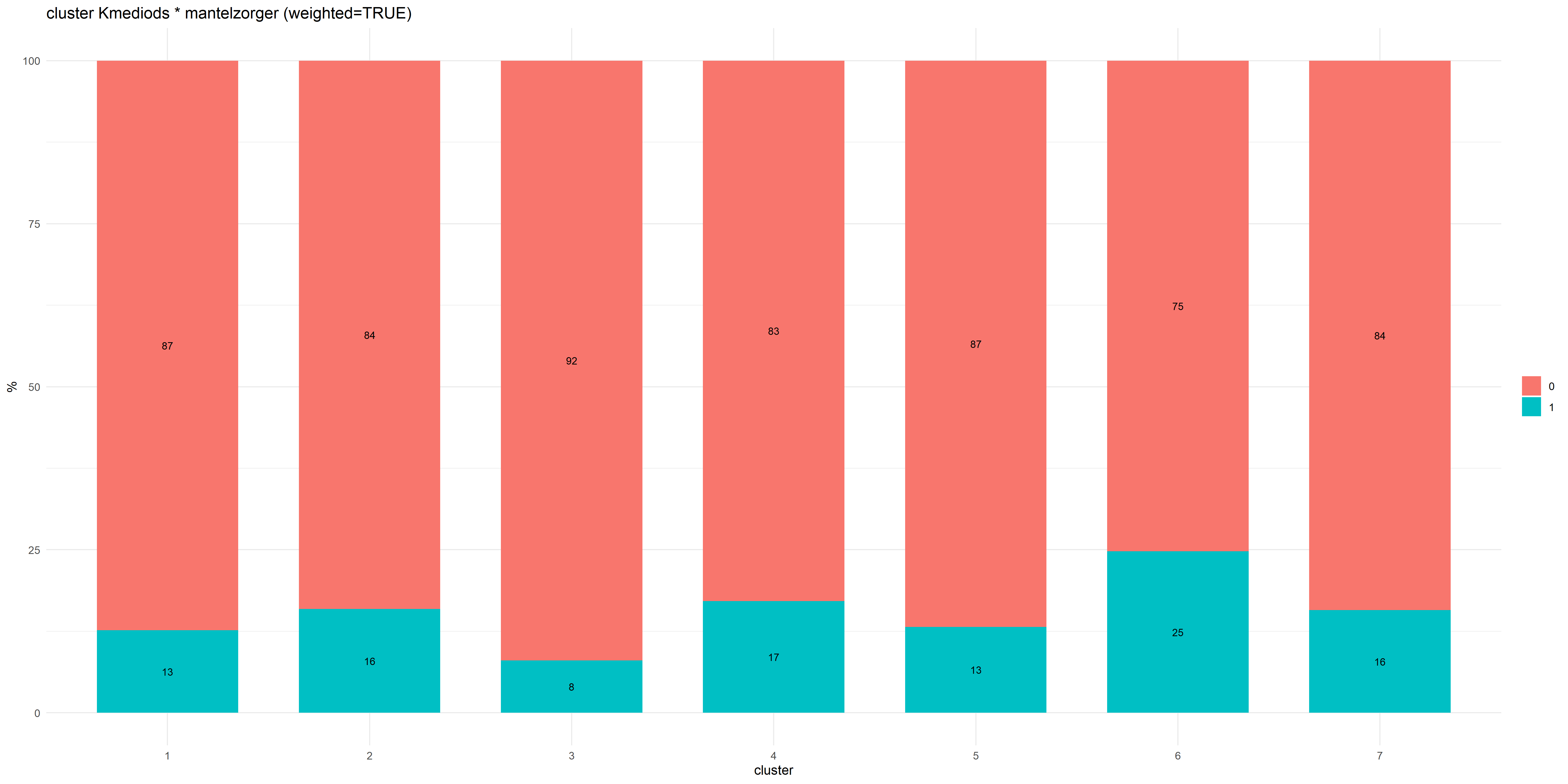
#### Vermoeidheid zonder duidelijke redenen (GGADB201)



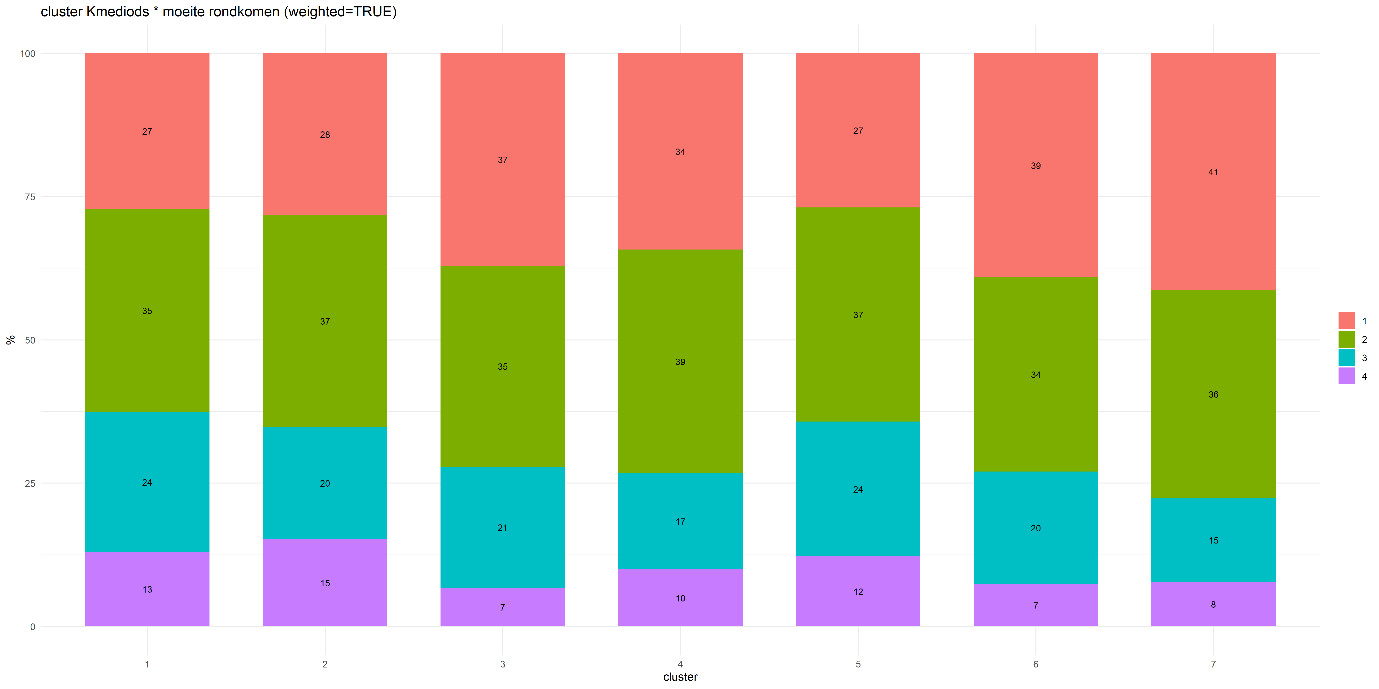
#### Mantelzorg ontvangen (MCMZOS304)



#### Mantelzorger (MCMZGS203)



#### Moeite met rondkomen (MMIKB201)



### Situationele / SES kenmerken

Alleenwonend

cluster 0 1

1 61.5 38.5

2 75.0 25.0

3 78.8 21.2

4 63.5 36.5

5 65.2 34.8

6 77.6 22.4

7 69.3 30.7

Eenoudergezin

cluster 0 1

1 94.20 5.80

2 90.79 9.21

3 92.65 7.35

4 97.12 2.88

5 91.80 8.20

6 91.37 8.63

7 94.67 5.33

Opleiding

cluster 1 2 3 4

1 25.41 37.67 23.05 13.87

2 9.39 20.62 33.81 36.18

3 4.62 15.83 39.34 40.20

4 13.69 39.17 26.81 20.32

5 13.12 30.68 28.31 27.90

6 6.28 24.86 29.84 39.02

7 12.79 41.35 21.94 23.92

Werkloos en werkzoekend

cluster 0 1

1 94.36 5.64

2 97.95 2.05

3 98.10 1.90

4 92.53 7.47

5 93.52 6.48

6 90.52 9.48

7 88.10 11.90

Pensioen

cluster 0 1

1 58.7268 41.2732

2 98.4994 1.5006

3 99.9763 0.0237

4 52.7680 47.2320

5 70.1626 29.8374

6 78.6174 21.3826

7 54.2080 45.7920

Huisvrouw/-man

cluster 0 1

1 77.525 22.475

2 97.517 2.483

3 99.616 0.384

4 72.631 27.369

5 85.436 14.564

6 83.050 16.950

7 64.350 35.650

Vrijwilligerswerk

cluster 1 2

1 17.4 82.6

2 26.5 73.5

3 13.4 86.6

4 32.3 67.7

5 22.1 77.9

6 68.1 31.9

7 21.2 78.8

Ten slotte is middels een Principal Component Analyse de samenloop van omstandigheden van groepen in de verschijningsvormen van kwetsbaarheid herleid. Daarvoor zijn de volgende situationele en SES-features gebruikt:

* leeftijd 70 en ouder (leeftijd70eo)
* Opleiding volwassenen laag / midden (opl\_lm)
* Betaald werk, opleiding afwezig (werkopleiding\_dich)
* Slechte gezondheid (ervarengezondheid\_dich)
* Langdurige ziekten (ziek\_lt)
* Mantelzorg ontvangen (MCMZOS304\_dich)
* Mantelzorg geven (mantelzorg\_dich)
* Vrijwilligerswerk (vrijwilligerswerk\_dich)
* Vriendenkring beperkt (vriendenkring\_dich)

Principal component analyse \* cluster

Loadings cluster 1:

RC1 RC4 RC2 RC3

leeftijd70eo 0.591

ervarengezondheid\_dich 0.760

werkopleiding\_dich 0.662

vrijwilligerswerk\_dich 0.635

opl\_lm 0.722

ziek\_lt 0.729

mantelzorg\_dich 0.509

MCMZOS304\_dich 0.706

vriendenkring\_dich 0.948

RC1 RC4 RC2 RC3

SS loadings 1.368 1.281 1.265 1.035

Proportion Var 0.152 0.142 0.141 0.115

Cumulative Var 0.152 0.294 0.435 0.550

Loadings cluster 2:

RC1 RC3 RC2 RC4

leeftijd70eo

ervarengezondheid\_dich 0.628

werkopleiding\_dich 0.827

vrijwilligerswerk\_dich 0.751

opl\_lm 0.744

ziek\_lt 0.806

mantelzorg\_dich 0.913

MCMZOS304\_dich 0.540

vriendenkring\_dich 0.602

RC1 RC3 RC2 RC4

SS loadings 1.946 1.230 1.090 1.030

Proportion Var 0.216 0.137 0.121 0.114

Cumulative Var 0.216 0.353 0.474 0.588

Loadings cluster 3:

RC1 RC2 RC3 RC4

leeftijd70eo 0.694

ervarengezondheid\_dich 0.582

werkopleiding\_dich 0.725

vrijwilligerswerk\_dich 0.703

opl\_lm 0.690

ziek\_lt 0.614

mantelzorg\_dich -0.550

MCMZOS304\_dich 0.566

vriendenkring\_dich 0.556

RC1 RC2 RC3 RC4

SS loadings 1.144 1.113 1.079 1.075

Proportion Var 0.127 0.124 0.120 0.119

Cumulative Var 0.127 0.251 0.371 0.490

Loadings cluster 4:

RC1 RC2 RC4 RC3

leeftijd70eo 0.613

ervarengezondheid\_dich 0.578

werkopleiding\_dich 0.872

vrijwilligerswerk\_dich 0.526

opl\_lm 0.484

ziek\_lt 0.503

mantelzorg\_dich 0.616

MCMZOS304\_dich 0.663

vriendenkring\_dich 0.855

RC1 RC2 RC4 RC3

SS loadings 1.362 1.161 1.115 1.054

Proportion Var 0.151 0.129 0.124 0.117

Cumulative Var 0.151 0.280 0.404 0.521

Loadings cluster 5:

RC1 RC2 RC4 RC3

leeftijd70eo 0.748

ervarengezondheid\_dich 0.756

werkopleiding\_dich 0.759

vrijwilligerswerk\_dich 0.544 0.473

opl\_lm

ziek\_lt 0.675

mantelzorg\_dich 0.762

MCMZOS304\_dich

vriendenkring\_dich 0.817

RC1 RC2 RC4 RC3

SS loadings 1.446 1.149 1.121 1.083

Proportion Var 0.161 0.128 0.125 0.120

Cumulative Var 0.161 0.288 0.413 0.533

Loadings cluster 6:

RC1 RC2 RC3 RC4

leeftijd70eo 0.675

ervarengezondheid\_dich 0.699

werkopleiding\_dich 0.835

vrijwilligerswerk\_dich 0.542 0.530

opl\_lm 0.611

ziek\_lt 0.685

mantelzorg\_dich 0.694

MCMZOS304\_dich -0.692

vriendenkring\_dich 0.874

RC1 RC2 RC3 RC4

SS loadings 1.835 1.173 1.050 1.012

Proportion Var 0.204 0.130 0.117 0.112

Cumulative Var 0.204 0.334 0.451 0.563

Loadings cluster 7:

RC1 RC4 RC2 RC3

leeftijd70eo 0.589

ervarengezondheid\_dich 0.946

vrijwilligerswerk\_dich 0.703

opl\_lm 0.757

ziek\_lt -0.661

mantelzorg\_dich 0.563

MCMZOS304\_dich 0.791

vriendenkring\_dich 0.638

RC1 RC4 RC2 RC3

SS loadings 1.383 1.194 1.150 1.025

Proportion Var 0.173 0.149 0.144 0.128

Cumulative Var 0.173 0.322 0.466 0.594

1. Omdat het algoritme overeenkomstige patronen in de ‘vingerafdruk’ van individuele kwetsbaarheid blootlegt, volstaat het om de prominente markers of features van het concept als uitgangspunt te nemen. M.a.w. het is niet nodig om het concept tot op de finesses af te dekken. [↑](#footnote-ref-1)
2. Ongestuurd bepalen van de natuurlijke grenzen in het antwoordbereik [↑](#footnote-ref-2)
3. https://stefvanbuuren.name [↑](#footnote-ref-3)
4. https://lvdmaaten.github.io/tsne/ [↑](#footnote-ref-4)
5. gewogen naar populatie zelfstandig wonende inwoners 19 jaar en ouder in de GGD regio Zuid-Holland Zuid. Cluster 0 omvat alle cases die niet zijn toegewezen aan een cluster uit de K-means analyse. [↑](#footnote-ref-5)