

**OPLEIDING: UNIVERSITEIT GENT, FACULTEIT WETENSCHAPPEN,  
OPLEIDING MASTER OF SCIENCE IN DE GEOGRAFIE EN DE GEOMATICA**

## **LESS IS MORE**

EVALUATIE VAN EEN NIEUWE PPGIS-METHODE OM  
GEOGRAFISCHE DATA TE VERZAMELEN

Aantal woorden: 14 548

Rien Boydens

Stamnummer: 01700429

Promotor: Prof. dr. Nico Van de Weghe

Copromotor: dr. Laure De Cock

Masterproef voorgelegd voor het behalen van de graad master in de richting  
Geografie en Geomatica

Academiejaar: 2022 - 2023

## WOORD VOORAF

Ik ben altijd al van het principe geweest om alles simpel te houden, *Keep It Stupidly Simple* of KISS. Toen professor Van de Weghe met het concept afkwam voor een tool waarbij mensen simpelweg iets kunnen tekenen op de kaart en dan nog wat korte vragen beantwoord – dit allemaal als Proof of Concept in PowerPoint – was ik wel geïnteresseerd. Niet alleen omdat het aansluit bij mijn principes van simpelheid, maar ook omdat ik hier een leuk programmeerproject in zag waarin ik mijn vaardigheden als webdeveloper kon verbeteren. Dit is ook gelukt, want ik kon de tools die ik hiervoor heb gemaakt gebruiken als voorbeeld in een jobinterview, en ik ben nu aangenomen als webdeveloper. Ik wist dat het programmeerproject maar een klein onderdeel ging zijn van deze masterproef en dat dit nog steeds een academisch onderzoek is. Desalniettemin was ik bereid om mij in te zetten om een goed onderzoek te leveren en mooie academische thesis te schrijven.

Ik zeg nu wel dat ik hiervoor bereid was, maar ik was ook bereid om heel veel andere zaken te doen. Ik ben soms wel een bezige bei, en houd me vaak met veel zaken tegelijkertijd bezig. Dit jaar was ik bijvoorbeeld naast het thesissen actief in drie verschillende studentenverenigingen waarvan één als voorzitter, nam ik nog een aantal extra vakken op met een kredietdoelcontract puur uit interesse, werkte ik één dag per week als jobstudent geomaticus, en was ik moderator van een online forum met meer dan vierduizend mensen en dan moest ik tussen al dit nog wat vrije tijd vinden om te ontspannen en uit te rusten. Hoe dat ik dit volhield, niet blijkbaar. De maand voor alle deadlines was het allemaal een beetje te veel geworden, en heb ik een paar zware burn-outs gehad. Toen wist ik dat ik het leven even wat rustiger moest nemen, en mij wat minder moet engageren in alles, maar mijn masterproef afwerken voor de deadline van eerste zit ging niet meer, zeker niet als ik er iets deftig van wou maken.

Toch ben ik nu blij dat ik eindelijk deze laatste woorden kan schrijven in dit document. Ik ben verheugd dat ik eindelijk mijn masterdiploma kan halen en een punt kan zetten achter mijn academische carrière. Dit alles was nooit mogelijk geweest zonder mijn ouders die mij heel de tijd bleven ondersteunen, of zonder alle vrienden die ik heb gemaakt in het Gentse die mij zo veel leut en lach hebben bezorgd. Vooral wil ik iedereen bedanken die mij begeleidde doorheen deze masterproef: mijn promotor professor dr. Nico Van de Weghe die zo geduldig met mij bleef wetende dat academisch lezen en schrijven niet mijn forte is, Sien Benoît mijn eerste begeider die mij inzichten gaf over haar eigen onderzoek rond de Fietsbarometer, dr. Laure de Cock mijn copromotor die mij altijd liet herinneren wat er belangrijk is in een academisch onderzoek, Lars De Sloover die mij op het einde nog hielp om dit werk nog af te krijgen, en Bart De Wit die mij zo veel heeft geholpen met de server om mijn webapplicaties op te zetten en het groot touchscreen met mijn vragenlijst dat we aan de inkom van de S8 hebben gezet. Ik vind het nog steeds zot dat doorheen de tijd dat ik aan deze masterproef werkte er al twee van mijn begeleiders de vakgroep hebben verlaten, Sien en Laure.

Rien Boydens  
25 juli 2023  
Nieuwpoort

## SAMENVATTING

Vaak is het interessant om de kennis van de alledaagse mens te bevragen om tot meer inzichten te komen en betere beleidsbeslissingen te nemen. Een Public Participatory GIS (PPGIS) onderzoek is een GIS onderzoek dat zich focust op de wijseden van de alledaagse mens. Het is dan ook belangrijk dat de gebruikte PPGIS applicatie gebruiksvriendelijk zijn voor leken, en dat de gebruikers niet te veel GIS concepten moet kennen om ermee om te gaan. In deze thesis onderzoeken wij een nieuwe methode voor PPGIS onderzoek, genaamd de Less Is More (LIM) methode, waarbij de focus ligt op de vragenlijst zo kort en simpel mogelijk te houden maar er toch nog veel inzichten uit te halen. Voor een case studie rond verkeersveiligheid hebben wij een online vragenlijst gemaakt bestaande uit vijf korte vragen. Deze bestaan vooral uit open vragen, aangevuld met een bevraging van geografische data die werd ingetekend aan de hand van een intuïtieve digitale markeerstift. Voor de analyse van deze verkregen data, worden de antwoorden op de open vragen gecategoriseerd aan de hand van Natural Language Processing (NLP) technieken, en is er een applicatie ontwikkeld om alles makkelijk en visueel te analyseren en interpreteren. We ondervonden dat de LIM-methode een sterke methode is om PPGIS onderzoek te doen, en dat we er veel kennis uit kunnen halen die anders totaal onbekend was voor de onderzoeker. Extra aandacht moet wel besteed worden aan het goed adverteren van de vragenlijst zodat er genoeg respondenten zijn. Bij de vergelijking met een gelijkaardig onderzoek aan de hand van een Quadrat Analysis met een  $\chi^2$  Goodness-of-Fit test, blijkt dat de LIM-methode ruimtelijke niet dezelfde resultaten krijgt als een uitgebreidere vragenlijst, maar de LIM-methode is wel intuïtiever en soms ook duidelijker. We sluiten af met nog suggesties waarvoor de LIM-methode ook nog gebruikt kan worden, zoals bij *crowd monitoring* en omwonende bevragingen.

## **ABSTRACT**

It is often interesting to gather the knowledge of the crowd to come to new insights and better policy making. A Public Participatory GIS (PPGIS) research is a GIS research that focuses on the wisdom of the crowds. It is then important that the used PPGIS application is user friendly for the layman, and that the users don't need a lot of GIS knowledge to operate the application. In this thesis we evaluate a novel method to do PPGIS research, called the Less Is More (LIM) method, that focuses on keeping the survey as short and simple as possible but also gaining a lot of insights. For a case study on traffic safety we developed an online survey of five short questions. These questions are mainly open-ended questions, supplemented by a survey of geographical data that gets drawn with an intuitive digital marker. To analyze the gathered data we categorized the answers to the open-ended questions with Natural Language Processing (NLP) techniques, and we developed an application to easily and visually analyze and interpret the data. We found that the LIM-method a strong method is to do PPGIS research, and that we can gain a lot of knowledge from it that was previously unknown to the researcher. Extra attention needs to paid to advertising the online survey so that we get enough participants. We compared the result of the case study with the results of a similar but a more comprehensive study with a Quadrat Analysis and a  $\chi^2$  Goodness-of-Fit test and found out that the LIM-method doesn't quite give the same results spatially. But the LIM-method is more intuitive and more clear in its results. We conclude our findings suggesting other applications for the LIM-method, like crowd monitoring and local resident surveys.

## INHOUDSTAFEL

Woord Vooraf.....	i
Samenvatting.....	iii
Abstract.....	iv
Inhoudstafel.....	v
1 Inleiding.....	1
1.1 Less Is More.....	3
1.2 Evolutie van de gebruiksvriendelijkheid van PPGIS-applicaties.....	6
1.3 Gebruiksvriendelijkheid in PPGIS applicaties.....	7
2 Data, Methoden en Studiegebied.....	9
2.1 Dataverzameling: de LIM-tool.....	9
2.2 Casestudie en studiegebieden.....	12
2.3 Data-Analyse.....	15
2.3.1 Open Vragen.....	15
2.3.2 Exploratieve analyse tool.....	19
2.3.2.1 Heatmap.....	21
2.3.2.2 Vervoersmodi.....	21
2.3.2.3 Color Map.....	21
2.3.2.4 Origin/Destination.....	21
2.3.2.5 Topics.....	22
2.3.2.6 Wordcloud & antwoorden.....	22
2.4 Vergelijking met de Fietsbarometer.....	23

3 Resultaten.....	25
3.1 Meta statistieken.....	25
3.2 Topics uit open vragen.....	28
3.2.1 Vraag 4: wat is er gevaarlijk aan.....	28
3.2.1.1 Evaluatie van de topics.....	33
3.2.2 Vraag 5: hoe zou u het verbeteren.....	38
3.2.2.1 Evaluatie van de topics.....	44
3.3 Voorbeelden uit de LIManalysis tool.....	48
3.3.1 Verkeersveiligheid voor zwakke weggebruikers.....	48
3.3.2 Veiligheid onderweg naar de S8.....	51
3.3.3 Houdt uw snelheid.....	52
3.3.4 Waar zijn er spiegels nodig.....	53
3.4 Vergelijking met de Fietsbarometer.....	55
3.4.1 Quadrant Analysis met een $\chi^2$ Goodness-Of-Fit test.....	55
3.4.2 Visuele vergelijking tussen LIM en de Fietsbarometer.....	57
4 Discussie.....	59
4.1 Dataverzameling.....	60
4.2 Data-Analyse.....	65
4.3 Andere gebruiken.....	69
5 Besluit.....	70
Referentielijst.....	73

## **1 INLEIDING**

In zijn boek *The wisdom of crowds* praat Surowiecki (2005) over hoe groepen betere beslissingen nemen dan individuen. Hij haalt hiervoor als voorbeeld een wedstrijd voor het raden van het gewicht van een rund op een dorpsfeest. Het gemiddelde van de gokken van alle mensen samen ligt dichter bij de waarheid dan de meeste gokken apart. Surowiecki probeert hierbij te bewijzen dat een divers publiek aan onafhankelijk beslissende individuen vaker betere beslissingen maakt dan enkele individuen of experten. De *crowd* heeft dus een uitgebreidere kennis van alle zaken dan één expert in zijn of haar vakgebied. In de Geografie is er al lang een notie van *the wisdom of crowds* in de vorm van Volunteerd Geographic Information (VGI) (Yan et al., 2020) en Public Participatory GIS (PPGIS) (Brown, 2012, 2012; Brown & Kyttä, 2014; Haklay & Tobón, 2003).

Bij de opkomst van Geografische Informatie Systemen (GIS) waren het vooral experten die kennis hadden van en toegang hadden tot deze krachtige tools voor het verzamelen, beheren en analyseren van geografische data. Deze geografische data werden ook vooral beheerd door overheidsinstanties en het publiek had er weinig toegang tot, maar met de opkomst van nieuwe technologieën en een groter belang aan burgerparticipatie, hebben de ideeën van Volunteerd Geographic Information en Participatory GIS aan populariteit gewonnen.

VGI verwijst naar geografische informatie dat wordt geleverd door individuele burgers en informele groepen (Goodchild, 2007). Het omvat het verzamelen, delen en beheren van geografische gegevens door mensen die geen professionele achtergrond hebben op het gebied van geografie of GIS. Deze gegevens kunnen variëren van locatiegegevens tot gedetailleerde beschrijvingen en foto's van plaatsen en gebeurtenissen. VGI wordt vaak verzameld via crowd-sourcing platforms, mobiele applicaties en sociale

media. PPGIS daarentegen is een bredere benadering die het gebruik van geografische informatiesystemen combineert met participatieve methoden om burgers te betrekken bij het nemen van beslissingen over ruimtelijke vraagstukken (R. Sieber, 2006). PPGIS omvat het verzamelen van gegevens door burgers, maar legt ook de nadruk op het betrekken van de lokale gemeenschap bij het identificeren van problemen, het verkennen van oplossingen en het nemen van beslissingen die van invloed zijn op hun leefomgeving. Het doel is om de kennis en ervaringen van de lokale bevolking te benutten en hun betrokkenheid bij ruimtelijke planning en besluitvorming te vergroten.

Het probleem met VGI en PPGIS is dat er toch nog basis GIS kennis van de mensen wordt verwacht, zoals de geografische entiteiten punten, lijnen en polygonen. Ook is het vaak veel werk om geografische informatie te creëren. Naast het intekenen van de ruimtelijke data, moet er ook veel attribuutdata ingevoerd worden. In onderzoek bestaat dit werk vaak uit het intekenen van een bepaald fenomeen door een punt te zetten, een route te tekenen of een gebied aan te duiden met een polygoon. Dan worden er nog een heel deel vragen gesteld over wat er is aangeduid wat veel tijd en moeite kost. Een voorbeeld van zo'n PPGIS-onderzoek is de Fietsbarometer (Storme et al., 2022). Daar worden de respondenten, scholieren, gevraagd om hun route in te tekenen van thuis naar school op basis van een bestaand wegennet, en dan een hele hoop variabelen te beoordelen die invloed hebben op waarom ze deze route nemen, en wat ze er gevaarlijk aan vinden.

In dit onderzoek proberen we een methode te ontwikkelen en evalueren, genaamd de Less Is More (LIM) methode, waarbij de nadruk ligt op zo veel mogelijk informatie te verzamelen met zo weinig mogelijk moeite van de participanten. We vragen of de LIM-methode een bruikbare methode is om PPGIS onderzoek te doen. Hiervoor kijken we of we interessante interpretaties kunnen halen uit de resultaten van de LIM-methode en kijken we of deze

interpretaties overeenkomen met bestaande onderzoeken. In de rest van sectie 1 leggen we de LIM-methode in meer detail uit, en halen we nog relevante onderwerpen aan die betrekking hebben tot de LIM-methode zoals PPGIS, het Web, en *Human-Computer Interaction*. In sectie 2 leggen we de methoden uit die we hebben gebruikt om de LIM-methode te onderzoeken. Hierbij spreken we over het verzamelen van de data met een Webgebaseerde tool, het analyseren van de data met Natural Language procesing (NLP) en een interactieve analyse tool, en hoe we het kunnen vergelijken met de resultaten van de Fietsbarometer, een uitgebreider onderzoek dat hetzelfde thema onderzoekt als onze casestudie, namelijk fietsveiligheid. Vervolgens bespreken we de resultaten in sectie 3. Dit bevat een verkennende meta-analyse over het aantal respondenten en wie deze respondenten zijn, vervolgens het proces van topics te extraheren aan de hand van NLP, een paar cases uit de casestudie om de analyse tool voor te stellen, en een kritische vergelijking met de resultaten van de Fietsbarometer. In sectie 4 volgt dan nog een discussie waar we een kritische kijk nemen op alles wat beter kon in de uitvoering van de casestudie, en een paar applicaties waarvoor de LIM-methode nog kan gebruikt worden.

## 1.1 Less Is More

Less Is More (LIM) bestaat erin om met zo weinig mogelijk, *less*, zo veel mogelijk te behalen, *more*. Zo weinig mogelijk houdt in zo weinig mogelijk moeite van de respondent, zo weinig mogelijk tijd dat de respondent moet spenderen aan het onderzoek, zo weinig mogelijk vragen dat de respondent moet oplossen, zo weinig mogelijk moeite van de onderzoekers om respondenten te krijgen, zo weinig mogelijk kosten om het onderzoek te doen, zo weinig mogelijk moeite doen om tot interpretaties en conclusies te komen. Met zo veel mogelijk bedoelen we zo veel mogelijk respondenten, zo veel mogelijk variatie in antwoorden dat het onderzoek krijgt, zo veel mogelijk informatie dat onderzoekers uit de antwoorden kunnen halen, zo veel

mogelijk antwoorden dat ruimtelijk geconcentreerd zijn zodat je er zinvolle conclusies uit kan trekken. Kortom, zo veel mogelijk halen uit zo weinig mogelijk.

Deze methode probeert zo een welbepaald probleem op te lossen in te PPGIS, namelijk dat onderzoeken vaak minder respondenten krijgen dan vroeger (Brown & Kyttä, 2014). Mensen willen niet meer zo vaak tijd vrijmaken voor lange vragenlijsten, of hebben de attentiespan niet meer om een lange bevraging af te werken. Bij bijvoorbeeld de Fietsbarometer (Storme et al., 2022) haken er een deel van de respondenten af. Van de 1256 respondenten dat ze hadden, hadden een kleine twintig procent de vragenlijst niet volledig ingevuld. Hierdoor konden deze antwoorden niet meer gebruikt worden. Ook waren de respondenten ruimtelijk sterk verspreid over verschillende steden, waardoor er vaak geen ruimtelijk geconcentreerd genoeg netwerk was om lokale kennis uit te halen. Een ander voorbeeld van een onderzoek waar er respondenten wegvielen is dit van Alattar et al. (2021). Zij onderzochten het verschil in actief transport tussen Strava gebruikers en niet-Strava gebruikers in Glasgow met een Web gebaseerde PPGIS-tool. Van de 1637 respondenten hebben er 1097 de vragenlijst volledig kunnen invullen, waarvan er 821 gebruikers actief transport gebruikten. Van de 531 routes van fietsers waren er 21 slecht getekend wegens problemen met de webapplicatie.

Vele Web gebaseerde PPGIS-bevragingen met willekeurige sampling hebben een participatie van 13% (Brown & Kyttä, 2014). Deze Web gebaseerde bevragingen werken dan vaak ook met een vooraf bepaalde steekproef groep, welke meestal aandeelhouders zoals omwonende zijn, waarnaar uitnodigingen gestuurd worden. Hierbij werd al geprobeerd om het aantal respondenten te verhogen met behulp van vergoedingen (Brown et al., 2014) of met gespecialiseerde internet panels (zie Brown & Kyttä, 2014). Dit zorgde maar voor een kleine verhoging van participanten of de kwaliteit van de data ging achteruit. Om betekenisvol PPGIS-onderzoek uit te voeren, moeten er ook

genoeg respondenten aan meedoelen. Jammer genoeg kampen vele PPGIS-onderzoeken met een globaal, over het volledige studiegebied (Brown et al., 2012), of lokaal, in één of meerdere deelgebieden (Storme et al., 2022), tekort aan respondenten (Brown & Kyttä, 2014) waardoor men er weinig informatie uit kan halen.

De meeste onderzoeken maken ook voornamelijk gebruik van gesloten vragen. Deze zijn makkelijker voor de respondent aangezien ze maar moeten klikken en niet de tijd moeten nemen om iets in te typen (Connor Desai & Reimers, 2019). Ook zijn deze makkelijker voor de onderzoeker om te analyseren wanneer ze online zijn ingevuld zonder toezicht van de onderzoeker. De respondent is dan wel gelimiteerd tot de antwoordmogelijkheden die de onderzoeker heeft gegeven. Onderzoekers proberen dit te limiteren door zo veel mogelijk antwoordmogelijkheden te geven, maar dan is er nog steeds een bias van de onderzoeker op de antwoordmogelijkheden en kan de lijst met mogelijke antwoorden lang oplopen. In tegenstelling heb je open vragen, waar de respondent alles op kan antwoorden waar die aan kan denken. Open vragen is dus een zeer sterk voorbeeld van de LIM-methode. Met één zeer gerichte open vraag, kan je zeer veel verschillende antwoorden verkrijgen zonder dat de respondent gelimiteerd is.

Met de LIM-methode proberen we dus met zo weinig mogelijk input van de onderzoeker en van de respondent, zo veel mogelijk output te krijgen van antwoorden en informatie. Hier zoeken we dus naar manier hoe dit mogelijk is. Een eerste al aangehaalde manier voor LIM is open vragen, waarbij je veel antwoordmogelijkheden geeft in één vraag. Een andere manier is om de vragenlijst aan te bieden via het internet. Bijna iedereen in deze hedendaagse maatschappij heeft een smartphone waarmee die eender welke webpagina kan bekijken en mee interageren. Door de vragenlijst via het internet aan te

bieden moeten respondenten niet veel moeite doen om deze te bekijken, en kunnen veel respondenten dit tegelijk invullen.

## **1.2 Evolutie van de gebruiksvriendelijkheid van PPGIS-applicaties**

Public Participatory GIS (PPGIS) is geëvolueerd over de jaren heen. Traditionele methoden om geografische data te verzamelen bij mensen waren het gebruik van een papieren kaart en een marker om zaken aan te duiden op die kaart (Pocewicz et al., 2012). Deze papieren kaarten kunnen later door een expert in een GIS-systeem gedigitaliseerd worden. Latere verbeteringen in computer hardware en software lieten toe om een GIS-systeem te bouwen dat minder technische kennis vroeg van de gebruiker. Hierdoor kon een leek digitaal zelf met het GIS aan de slag, alhoewel vaak nog een expert aanwezig moest zijn om het systeem uit te leggen en om bijstand te bieden (Haklay & Tobón, 2003). Een nieuwe focus in het PPGIS-onderzoek begon te liggen op Human Computer Interaction (HCI), waardoor GIS-systemen steeds toegankelijker werden (Haklay & Tobón, 2003).

De opkomst van het Web 2.0 en het Geospatial Web (Corbett & Cochrane, 2020) zorgde voor veel meer applicaties en mogelijkheden die cartografie en GIS makkelijk maken voor het brede publiek. Het GeoWeb is een collectie van online GIS-technologieën die het mogelijk maken om online geografische data te maken, bekijken en beheren (R. E. Sieber et al., 2016). Zhang (2019) deelt publieke participatie in het GeoWeb op in twee categorieën: participatie dat het GeoWeb gebruikt, en participatie dat intrinsiek is aan het GeoWeb. Voorbeelden zijn P/PGIS, Geo-questionnaires, begeleide VGI, en Neocartografie voor de eerste categorie; en open data zoals OpenStreetMap, en passieve VGI voor de tweede categorie. Geo-questionnaires (Czepkiewicz et al., 2018; Jankowski et al., 2016) is het soort PPGIS op het GeoWeb dat dit onderzoek gebruikt, en bestaat uit online enquêtes waarbij er ook naar ruimtelijke data gevraagd wordt.

Het grote probleem bij de meeste PPGIS applicaties, is dat de gebruiker nog steeds basis GIS kennis nodig heeft van geografische entiteiten zoals punten en polygonen (Haklay & Tobón, 2003). Zeker op mobiele toestellen zijn de klassieke inputmogelijkheden van punten en polygonen moeilijker te gebruiken. Ramírez Aranda et al. (2021) stellen hiervoor een nieuwe geografische entiteit voor, de digitale marker. Deze digitale marker werkt net zoals de vooraf beschreven fysieke marker op een papieren kaart waarmee een gebruiker intuïtief een gebied kan aanduiden op de kaart. Deze digitale marker is ook handiger op mobiele toestellen aangezien de vinger als stift dient die tekent op het kaartblad van het touchscreen. Dit kan een grote vooruitgang geven in gebruikerservaring van PPGIS applicaties.

### **1.3 Gebruiksvriendelijkheid in PPGIS applicaties**

De belangrijkste componenten in een PPGIS onderzoek zijn de gebruikers of participanten van dit onderzoek. Het is de inbreng van de participanten dat het onderzoek doet werken, en daarom moet er bij een PPGIS onderzoek aandacht besteed worden aan de gebruiksvriendelijk van de gebruikte applicaties voor de gebruikers (Haklay & Tobón, 2003). Het onderzoeksfield dat zich bezig houdt met het bestuderen van hoe mensen omgaan met computerapplicaties en hoe deze gebruiksvriendelijk kunnen zijn is Human Computer Interaction (HCI) (Sinha et al., 2010). In HCI zeggen ze dat een gebruiksvriendelijke applicatie niet de focus moet leggen op de onderliggende architectuur en concepten van de applicatie, maar wel de focus moet leggen op het werk dat de gebruiker met de applicatie uitvoert en de stappen dat de gebruiker moet volgen om dit werk te doen. In een PPGIS applicatie houdt dit dus niet in dat de gebruiker moet weten wat een GIS is en hoe deze in elkaar zit om met de applicatie te werken, maar juist moet weten wat er verwacht wordt dat deze gebruiker doet (Haklay & Tobón, 2003).

Om te testen of een applicatie gebruiksvriendelijk is voor de gewenste gebruikers, wordt er vaak een *Usability Evaluation* (Cockton, 2012) gedaan. Een *Usability Evaluation* beoordeelt of een applicatie of interactief systeem intuïtief is en makkelijk te gebruiken. Haklay & Tobón (2003) bekeken de synergie tussen PPGIS en HCI aan de hand van *Usability Evaluations* in drie workshops. Zo vonden vier kernzaken die helpen bij het ontwikkelen van gebruiksvriendelijke PPGIS applicaties. Ten eerste is het belangrijk om de gebruikers of participanten goed te informeren over geplande aanpassingen in hun omgeving en wat er van hun verwacht wordt. Ten tweede moeten de GIS lagen goed gepresenteerd worden. Gebruikers vonden het interessant en gemakkelijk om toegang te hebben tot zowel luchtfoto's als een goede basiskaart om hun op te oriënteren. Ten derde moet het gebruik van een kaart zo simpel mogelijk zijn zodat operaties zoals zoomen en de kaart verplaatsen intuïtief zijn. Ten vierde moet de volledige applicatie makkelijk te gebruiken zijn zodat de gebruik goed weet dat een actie gelukt is en genoeg feedback krijgt over wat er gebeurt.

Hier halen we nog een paar voorbeelden aan van welke ontwerpbeslissingen verschillende auteurs hebben genomen om hun PPGIS applicaties gebruiksvriendelijk te maken. Bugs et al. (2010) creëerden een Web 2.0 PPGIS applicatie waarbij gebruikers commentaar konden geven op geplande projecten in het straatbeeld van Canela, Brazilië. Gebruikers konden hier verschillende GIS-lagen aan informatie raadplegen en commentaar achterlaten op een plaats met de comment-tool. Hun *Usability Evaluation* toonde dat gebruikers dit makkelijk te gebruiken vonden en zeer informatief, maar bijna de helft van hun gebruikers had ervaring met een GIS en vonden dat andere mensen dit misschien niet zouden gebruiken. ChangeExplorer (Wilson et al., 2019) is een applicatie voor smart wearables waar gebruikers een notificatie krijgen op hun smartwatch als ze in een gebied komen waar er geplande veranderingen zijn. De gebruikers kunnen dan aan de hand van een paar simpele vragen commentaar geven op de plannen. Of ze kunnen

spontaan commentaar geven over een gebied waar ze veranderingen willen zien. De gebruikers en stadplanners waren er zeer positief over, maar de gebruikers vonden wel dat je te gehaast commentaar moest geven. Ook wandelen nog niet veel mensen rond met smart wearables.

Vele auteurs bouwen hun PPGIS applicatie met een User Centered Design aanpak (UCD) (ISO, 2009). UCD beschrijft een vroege en actieve focus op de noden van de gebruiker met een klemtouw op het iteratief verbeteren van de gebruikersinterface. Roth et al. (2015) gebruikten UCD bij het ontwikkelen van GeoVISTA CrimeViz, een webapplicatie voor het visueel analyseren van misdaad in tijd en ruimte. Alhoewel dit niet zozeer een PPGIS applicatie is, wordt het nog steeds gebruikt door mensen zonder GIS kennis. Met deze UCD aanpak hadden ze al snel door dat ze verschillende gebruikersprofielen hebben en dat dit moet gereflecteerd worden in de finale applicatie. Newman et al. (2010) gebruikten een late UCD aanpak bij het ontwikkelen van hun citizen science applicatie om invasieve soorten te inventariseren. Ze deden pas gebruikerstesten als de applicatie af was en ondervonden dat er zeer weinig functionaliteit intuïtief was voor de gewone gebruiker. Er waren veel te veel GIS specifieke zaken dat gebruikers niet verstonden.

## **2 DATA, METHODEN EN STUDIEGEBIED**

### **2.1 Dataverzameling: de LIM-tool**

De LIM-methode bestaat er uit om zo veel mogelijk data te verzamelen met zo weinig mogelijk moeite van de respondenten. Daarom is er gekeken naar vragen waarbij de respondent veel informatie kan voorzien met één antwoord. Eén zo'n mogelijkheid is open vragen. Open vragen laten toe dat respondenten antwoorden geven dat vooraf onbekend was door de onderzoeker. Een ander voordeel aan open vragen is dat de onderzoeker niet onbewust de antwoorden van de respondenten beïnvloed (Züll, 2016). Een

nadeel aan open vragen tegenover gesloten vragen is dat het moeilijker is om de antwoorden te analyseren (Popping, 2015).

Een andere mogelijkheid is om de respondenten het makkelijk te maken om veel informatie te geven, ligt bij het verzamelen van ruimtelijke data. Ramírez Aranda et al. (2021) vonden dat een digitale markeerstift beter de ruimtelijke realiteit van de respondent weergeeft en gebruiksvriendelijker is dan lijnen of polygonen tekenen zoals klassiek in de GIS wordt gedaan. Deze digitale marker werkt net zoals een fysieke marker op een papieren kaart waarmee een gebruiker intuïtief een gebied kan aanduiden op de kaart. Deze digitale marker is ook handiger op mobiele toestellen aangezien de vinger als stift dient die tekent op het kaartblad van het touchscreen. Het zorgt voor vloeiende lijnen die zeer makkelijk aan te duiden zijn met touchscreen toestellen zoals smartphones.

Aangezien er geen bestaande web-survey applicaties zijn die ons toelaten om een interactieve kaart toe te voegen met een aangepaste digitale marker, is voor het verzamelen van de data is een eigen webapplicatie ontwikkeld. Deze webapplicatie omvat een korte vragenlijst van 5 simpele vragen. De vragen worden één per één gepresenteerd aan de respondent en bestaan uit één meerkeuzevraag, twee vragen waar men iets moet aanduiden op een kaart, en twee open vragen waarvan er maar één verplicht is. Een overzicht van de vragen is te vinden in Tabel 1.

**Tabel 1: Vragen LIM-tool**

Vraag	Type	Opmerkingen
1 Hoe verplaatst u zich over dit verkeerspunt? Meerkeuze		Niet-elektrische fiets, Elektrische fiets, Speedpedelec, Niet-elektrische buitenmaatse

fiets,

Elektrische  
buitenmaatse fiets,  
Skate, Te voet,  
Bromfiets, Motor, Auto,  
Bus, Vrachtwagen,  
Andere

- 2 Teken de route die u volgt als u zich over dit kruispunt verplaats met uw gekozen vervoersmiddel.
- 3 Duid één plaats aan waar op dit kruispunt u het gevaarlijk vindt met uw gekozen vervoersmiddel.
- 4 Waarom vindt u het hier gevaarlijk? Open vraag Ten minste vijf woorden.
- 5 Hoe zou u dit verbeteren? Open vraag Optioneel.

---

In de eerste vraag wordt er gepeild naar het dominante vervoersmiddel van een gebruiker als hij/zij zich/haar over dit verkeerspunt rijdt. Zo kan er ook een opdeling gemaakt worden per vervoersmiddel bij de data-analyse. Aan de gebruiker wordt dan gevraagd om de volgende vragen in te vullen voor dit gekozen vervoersmiddel.

De tweede en derde vraag gebruiken een digitale marker zoals beschreven door Ramírez Aranda *et al.* (2021). Voor de tweede vraag is een dunne marker gebruikt om makkelijker lijnen te tekenen voor de routes. Met de getekende routes kunnen inzichten bekomen worden over de verkeersflows die over het

verkeerspunt passeren en kunnen inzendingen gefilterd worden op oorsprong en destinatie. Bij de derde vraag is de marker iets dikker om gebieden mee aan te duiden die mensen gevaarlijk vinden.

De vierde en vijfde vraag zijn open vragen. Een goede open vraag moet kort zijn en gesteld worden met het correcte vraagwoord (Popping, 2015), maar moet ook de respondent motiveren om de vraag in te vullen (Smyth et al., 2009; Züll, 2016). Bij web gebaseerde enquêtes is het belangrijk dat de grootte van het tekstveld overeenstemt met de lengte van het verwachte antwoord (Smyth et al., 2009; Züll, 2016), daarom worden de open vragen gepresenteerd met meerdere regels voor antwoordmogelijkheden waarbij de regels onderlijnd zijn alsof het op papier gepresenteerd wordt.

Bij het indienen van de vragenlijst krijgen de respondenten de optie om nog een route te tekenen samen met nieuw vervoersmodus en gevarengebied. Dit betekent dus de volledige vragenlijst opnieuw invullen voor hetzelfde studiegebied. Of ze kunnen ook een extra gevarengebied aanduiden voor de route die ze al hebben getekend. Dan wordt de vervoersmodus van vraag 1 en route van vraag 2 overgenomen van hun vorige inzending. Ook wordt bij het indienen bijgehouden hoe lang ze over de vragenlijst hebben gedaan, hoe ze bij de vragenlijst zijn gekomen, en de User-Agent waarmee ze de vragenlijst hebben ingevuld om te zien of ze via desktop of mobile dit invulde.

## **2.2 Casestudie en studiegebieden**

Als casestudie vergelijken we de *Less Is More*-methode met een bestaand PPGIS-onderzoek, namelijk de Fietsbarometer van Storme et al. (2022). Om de vergelijking te kunnen maken wordt hetzelfde thema onderzocht, namelijk verkeersveiligheid. Hierbij worden er een aantal factoren bekeken en vergeleken met de resultaten van de Fietsbarometer. Deze factoren zijn het aantal respondenten en de ruimtelijke densiteit van de inzendingen, de overeenkomsten en de verschillen van de resultaten tussen de LIM-methode

en de Fietsbarometer, de bruikbaarheid van de resultaten voor beleidsbeslissingen, en of de nieuwe methode de bestaande methode kan verrijken met nieuwe inzichten.

Om mensen te leiden naar de vragenlijst, worden posters met QR-codes opgehangen rond het studiegebied. Deze posters kunnen variëren van gewone posters of flyers op een willekeurige plaatse, tot posters bij handelszaken en flyers bij bus- en tramhaltes. Om te weten welke van deze adverteermethoden het beste werkt, wordt het studiegebied en de adverteermethode mee gecodeerd in de URL op de QR-code.

Als onderzoeksgebied is het rondpunt van De Sterre genomen in Gent (Figuur 1). Dit verkeerspunt is gekozen omdat er hierover veel data beschikbaar is in de Fietsbarometer, en omdat de onderzoekers hier heel bekend mee zijn. Het rondpunt De Sterre is een achtvormig rondpunt waar er vijf wegen op uitkomen: tweemaal de Kortrijksesteenweg, de Oudenaardsesteenweg, de Krijgslaan, en de Voskenslaan. Onder het dunne midden van de achtvorm loopt er een ruime fietserstunnel. Aan de noordkant van het rondpunt bevindt er zich een Delhaize supermarkt en een JIMS fitness. Aan de zuidkant bevindt er zich een tankstation en de universiteitscampus De Sterre. Deze campus kan betreden worden van het rondpunt via een wandelpaadje ten oosten van het tankstation of een doorgang met hekjes ten westen van het station. Bushaltes bevinden zich in de Oudenaardsesteenweg en de Kortrijksesteenweg in het westen. Een kilometer ten noorden van het rondpunt bevindt zich het treinstation Gent Sint-Pieters. Deze is te bereiken via de Voskenslaan of via de Kortrijksesteenweg. Ten westen van het rondpunt bevinden zich de buitenring van Gent (R4) en de autosnelweg (E40). Deze zijn te bereiken via de Oudenaardsesteenweg en de Kortrijksesteenweg respectievelijk.



**Figuur 1: Studiegebied De Sterre**

Reclame maken aan de hand van flyers en posters ging hier moeilijk aangezien maar weinig handelszaken deze wouden ophangen. Het was vooral op bushokjes dat deze gingen. De universiteitscampus De Sterre waar de vakgroep geografie zit, is wel vlakbij. Dus werd aan de studenten en andere onderzoekers gevraagd om de vragenlijst eens in te vullen en te beoordelen. Ook stond een week lang een groot touchscreen aan de ingang van het gebouw S8, waar de vragenlijst op gepresenteerd werd zodat mensen die passeerden dit konden invullen. Er was gepland om ook andere gebieden te onderzoeken, maar hiervoor is geen tijd gevonden.

## **2.3 Data-Analyse**

Om een algemeen overzicht te krijgen over de verzamelde data, worden eerst een paar meta-statistieken bepaald zoals aantal participanten, verdeling vervoersmiddel, of de vragenlijst op mobile of desktop is ingevuld, hoe de participant bij de vragenlijst is gekomen, en hoe lang ze over de vragenlijst deden. Zo kunnen al inzichten verworven worden over de effectiviteit van de vragenlijst. Hiermee beantwoorden we vragen zoals: hebben we veel participanten kunnen aantrekken, welk vervoermiddel is er het meest beoordeeld, is het vooral op mobile ingevuld waarvoor het is ontworpen, welke manier van reclame maken werkt het best, is het kort en simpel genoeg dat het binnen de 3 minuten kan ingevuld worden?

### **2.3.1 Open Vragen**

Om de open vragen, met open antwoorden efficiënt te kunnen gebruiken, moeten de antwoorden discreet worden. De open vragen moeten dus gesloten worden voor de analyse. Dit wordt het informatie-paradigma genoemd (Popping, 2015). Hier wordt dit gedaan door de antwoorden te categoriseren volgens onderwerp. Zo kan per locatie in een oogopslag gekeken worden waarover er wordt gesproken. Eén manier om dit te doen is door de antwoorden te coderen (Popping, 2015; Züll, 2016). Een set van categorieën wordt hierbij vooraf opgesteld, en een menselijke codeerdeerder of computer overloopt alle antwoorden en steekt deze in één van deze categorieën. Hiervoor moeten er echter vooraf gedefinieerde categorieën zijn, wat niet het geval is. Wij willen de verborgen onderwerpen eruit halen en ze zo classificeren. Ook is deze codering zeer werkintensief.

Beter methoden om deze antwoorden te classificeren komen uit de studie van Natural Language procesing (NLP). NLP is een verzameling van methoden om menselijke taal beschikbaar te stellen voor computers (Eisenstein, 2019). Dit wordt gebruikt voor lexicografische analyses, tekst voorspelling,

spraakherkenning, en voor ons geval classificatie van verschillende stukken tekst. In NLP worden alleenstaande stukken tekst, zoals antwoorden op vragen, documenten genoemd. Deze documenten bestaan uit één of meerdere zinnen, en deze zinnen zijn opgebouwd uit tokens. Tokens kunnen woorden zijn, maar ook delen van woorden, woordgroepen en leestekens. Een verzameling van  $N$  tokens noemt men een  $N$ -gram. Zo is een verzameling van twee token een bigram, en van drie tokens een trigram. De volledige verzameling van documenten wordt het corpus genoemd.

Voor tekstclassificatie wordt er in NLP vaak gebruik gemaakt van Bayesiaanse methoden (Van Loon, 2022). Veelgebruikte tekstclassificatie methoden zijn: Latent Dirichlet Allocation (LDA) (Blei et al., 2003), een generatief probabilistisch model waarbij elk document wordt gemodelleerd als een collectie van onderliggende topics; Latent Semantic Indexing (LSI) (Deerwester et al., 1990), een methode voor het automatisch indexeren en ophalen van documenten gebruikmakend van de hogere structuren van de documenten; en Non-Negative Matrix Factorization (NMF) (Wang & Zhang, 2013), een dimensiereductie paradigma dat een matrix probeert op te delen in kleinere matrices met niet-negatieve componenten en er zo verborgen structuren uitgehaald.

Al deze methoden maken gebruik van een bag-of-words representatie van de documenten (Ciganaitye et al., 2014; Jáñez-Martino et al., 2023). Dit is numerieke methode waarbij de frequentie van woorden in een document door een vector wordt weergegeven. Om deze bag-of-words representatie te bekomen is het eerst belangrijk om de data eerst op te kuisen. Opkuisen bestaat uit het wegdoen van veelvoorkomende stopwoorden zoals "en" en "ik", vervolgens het basiswoord of lemma nemen van de woorden, en woorden die maar in één document voorkomen wegfilteren. Met de overgebleven woorden wordt een dictionary opgebouwd dat aan elk uniek woord een nummer geeft. Dit nummer komt dan overeen met de dimensie in

de vector. Deze vector wordt dan per document gevormd door alle frequenties van woorden in dat document te tellen en te normaliseren. Vaak wordt ook de Term Frequency – Inverse Document Frequency (Tf-Idf) transformatie van een bag-of-words gebruikt. Hierbij wordt de frequentie van een token gedeeld door het aantal documenten in het corpus waar dit token voorkomt, met als doel om jargon specifieke stopwoorden weg te filteren. Voor het aanmaken van een bag-of-words gebruiken wij tokens bestaande uit N-grammen waarbij  $N \in \{1, 2, 3\}$ .

Wij hebben gemerkt dat de klassieke NLP-methoden voor topic modelling (LDA, LSI, NMF) niet goed werkten op onze verkregen data. Dit omdat het eerst al moeilijk om het aantal topics op voorhand in te schatten, welke bij alle drie nodig is, maar ook omdat de statistische eigenschappen van de woorden niet voldoen aan de verwachtingen van deze bayesiaanse methoden, zoals te zien zal zijn in sectie 3.2, zijn er zeer verschillende manieren om hetzelfde topic aan te halen. Daarom wordt er gekozen voor een hiërarchische clustering (HC) van de bags-of-words (Steinbach et al., 2000). Dit laat ons toe om op basis van een visuele vertoning van de afstanden tussen de bags-of-words het aantal clusters te kiezen.

Als afstandsmaat wordt de cosinus dissimilariteit (Lahitani et al., 2016) gebruikt. De cosinus similariteit is een maat die de richtingen van vectoren vergelijkt met elkaar en de hoekafstand tussen deze richtingen beschrijft. De formule van de cosinus similariteit wordt afgeleid van het euclidische inwendig product van de twee vectoren.  $\text{cosine similarity} = \frac{A \cdot B}{\|A\| \|B\|}$ . De cosinus dissimilariteit is dan één min de cosinus similariteit. Als koppelingsmethode (*linkage*) vonden we dat de Ward methode (Ward, 1963), welke de totale intra-cluster variantie zo klein mogelijk houdt, het beste bij onze data past. Ward gaat er van uit dat een Euclidische afstandsmaat gebruikt wordt, wat normaal niet het geval is bij de cosinus similariteit, maar de bag-of-words

representatie en de Tf-Idf variant stelt een document voor in een multidimensionale, L2-genormaliseerde vectorruimte. Omdat de vectorruimte L2-genormaliseerd is, is de cosinus similariteit proportioneel evenredig met de Euclidische afstand tussen de vectoren (Ross, 2020). Alhoewel dit geen formeel bewijs is, vinden vele auteurs dit goed genoeg om Ward's linkage samen te gebruiken met de cosinus dissimilariteit in de context van NLP (Altuncu et al., 2020; Amine et al., 2010; Jáñez-Martino et al., 2023; Sun & Korhonen, 2011). Ze vonden ook dat Ward met de cosinus dissimilariteit beter presteert dan andere linkage methoden zoals *single*, *average* en *complete*.

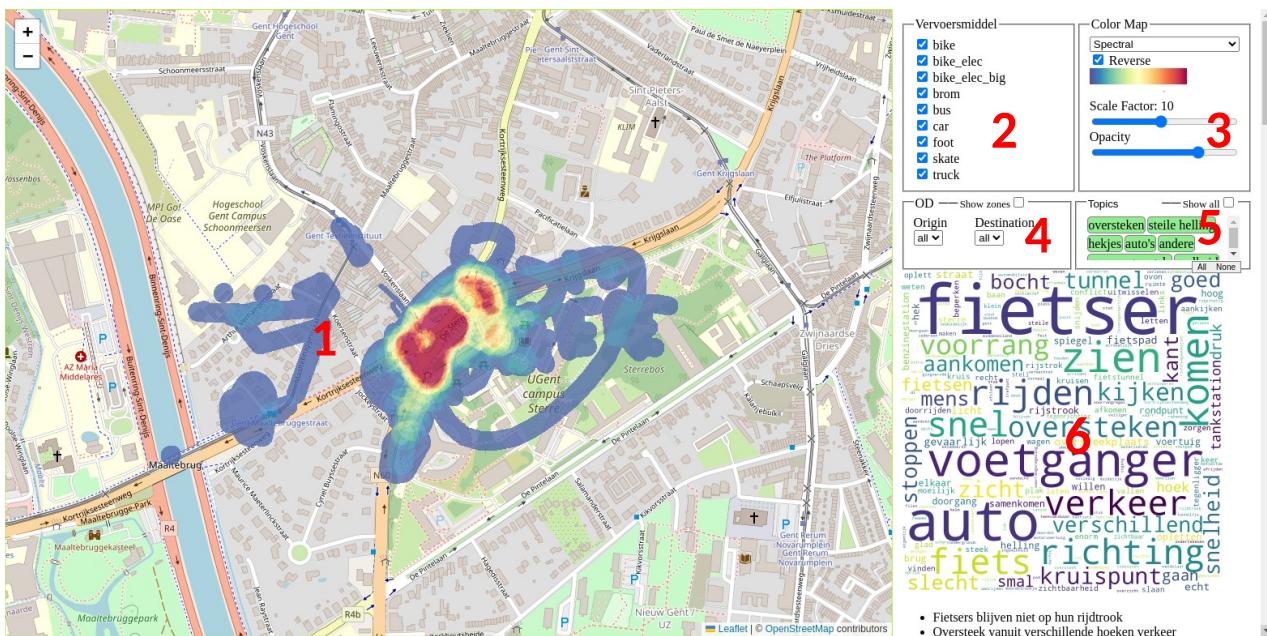
De clustering wordt gedaan met de implementatie van SciPy (Virtanen et al., 2020). Voor het plotten van het dendrogram biedt SciPy automatische inkleuring van de clusters op basis van de threshold  $0.7 * \max(\text{linkage threshold})$ . Visueel wordt gezocht naar een ideale indeling van de clusters op basis van het dendrogram en een heatmap van de paarsgewijze afstanden van de documenten. Elke cluster wordt vervolgens bekeken en een categorienaam wordt aan de cluster gegeven gebaseerd op de documenten in die cluster. Vaak is er dan nog een cluster "andere" die nog uit te veel verschillende topics bestaan. Dan wordt de clustering herhaald op alle documenten met de categorie "andere" om een fijnere clustering te bekomen. Door zo iteratief te werken worden clusters die duidelijk te identificeren zijn met een grovere clustering, niet gesplitst door een fijnere clustering.

Als we alle topics hebben gevonden, zullen we deze nog evalueren aan de hand van Multidimensional Scaling (MDS) (Hout et al., 2013). MDS bestaat uit een reeks statistische technieken om de dimensionaliteit en de complexiteit van data te verminderen zodat deze beter visueel geïnterpreteerd kan worden. Concreet houdt dit in dat we een multidimensionale ruimte, zoals onze document vectoren, zullen presenteren op een 2-dimensioneel vlak, waarbij de onderlinge afstanden zo goed mogelijk behouden blijven. Dit laat ons toe om nog eens visueel te inspecteren of onze topics in een 2-

dimensionele ruimte weldegelijk dichte clusters vormen. Ook worden de verkregen clusters vergeleken met de latente topics verkregen met LDA en NMF. Ook al zijn de topics die we verkrijgen met LDA en NMF niet goed te identificeren en interpreteren, helpen deze vergelijkingen toch om latente verbanden tussen topics zichtbaar te stellen. HC-topics die uit dezelfde latente topics bestaan, kunnen bijvoorbeeld verbonden zijn met elkaar. Deze topic extractie wordt zowel gedaan voor de antwoorden op Vraag 4, die vraagt achter de reden waarom het gevaarlijk is, en de antwoorden op Vraag 5, die vraagt achter verbeteringen.

### **2.3.2 Exploratieve analyse tool**

De bedoeling van de LIM methode is ook om het voor een domeinspecialist makkelijk te maken om de verkregen data te analyseren en interpreteren. Daarom is er voor een makkelijke en exploratieve analyse van de resultaten, een interactieve tool ontwikkeld, genaamd LIManalysis (Figuur 2), dat een onderzoeker toelaat om verschillende aspecten van de resultaten te visualiseren en inspecteren en deze te filteren op bepaalde criteria. Onze tool, zoals te zien in , toont een heatmap van alle zones die zijn aangeduid als gevaarlijk, alsook een wordcloud van de schoongemaakte woorden in de antwoorden op Vraag 4 of Vraag 5, en eronder de antwoorden zelf. Het laat ook toe om te filteren op vervoermodus, oorsprong en destinatie van de route, en gedetecteerde topics. De tool is geschreven als een Webapplicatie van de Model/View/Controller architectuur (Leff & Rayfield, 2001). Het Model is de opgekruiste data in de databank op de server dat geserveerd wordt met Python en Flask voor de backend. De View is de heatmap, wordcloud en antwoorden, en de Controller zijn de verschillende filters die kunnen toegepast worden. De View en Controller zijn geschreven in HTML, CSS en JavaScript en is wat de onderzoeker te zien krijgt.



**Figuur 2: Limanalysis tool**

### 2.3.2.1 Heatmap

De heatmap is de voornaamste visualisatie van de data. Hierop kan de onderzoeker in een oogopslag zien waar mensen het gevaarlijkst vinden gebaseerd op de aangeduide filters. De onderzoeker kan ook op een bepaalde plaats klikken om info te zien, in de topics, wordcloud en antwoorden, voor juist die plaats.

### 2.3.2.2 Vervoersmodi

Een eerste filter is de vervoersmodus die werd ingegeven. Met deze filter kan makkelijk gezien worden waar het gevaarlijk is voor bijvoorbeeld zwakke weggebruikers, voor fietsers, voor gemotoriseerd verkeer, etc.

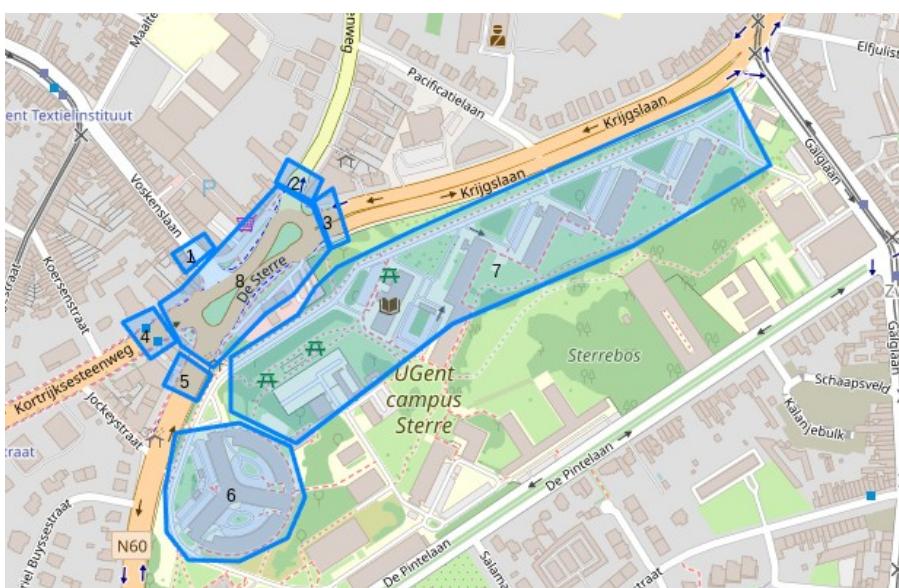
### 2.3.2.3 Color Map

De kleurenmap waarmee de heatmap getoond wordt kan aangepast worden voor een duidelijker visualisatie. De belangrijkste parameter hierbij is de

Scale Factor. Deze bepaald hoe hoog de kleurenmap geschaald wordt. In het voorbeeld betekent dit dat bij een grotere schaalfactor, er meer rood zal zijn. De andere parameters zijn de gebruikte kleurenmap zelf, en de opaciteit waarmee de heatmap wordt getoond om de onderliggende kaart nog te zien.

#### 2.3.2.4 Origin/Destination

De origin en destination zijn berekend op basis van de ingetekende routes. Hiervoor zijn er acht OD-zones bepaald door de onderzoeker, zoals in Figuur 3. De origin is bepaald als de zone waardoor de ingetekende route als eerste gaat. De destination is bepaald als de zone waardoor de ingetekende route als laatste gaat. De onderzoeker kan dan filteren op de origin en destination van de route waarbij de gevarenzone hoort.



**Figuur 3: Origin en Destination zones**

#### 2.3.2.5 Topics

De topics, zoals bepaald in 2.3.1, zijn zeer interessant om op te filteren. De onderzoeker kan de topics selecteren waarop dan ook wordt gefilterd voor de berekening van de heatmap en het creëren van de wordcloud en het tonen

van de antwoorden zelf. Ook zijn enkel de topics zichtbaar die voldoen aan de andere gegeven filters en geklikte positie op de heatmap, tenzij “Show all” aangeduid is. Naast de topics van Vraag 4, kunnen ook de topics over de verbetering van Vraag 5 gebruikt worden.

#### 2.3.2.6 Wordcloud & antwoorden

De wordcloud en antwoorden zijn geen filters. De wordcloud toont alle woorden, opgekuist en gelemmatiseerd, en de frequentie van deze woorden die zitten in de antwoorden waarop de filters van toepassing zijn. Wordclouds zijn een populaire en efficiënte methode om tekst samen te vatten en in een oogopslag te tonen waar het over gaat (Heimerl et al., 2014). Onder de wordcloud zijn ook nog de individuele antwoorden te lezen.

### 2.4 Vergelijking met de Fietsbarometer

Om te zien of de LIM methode dezelfde resultaten geeft als de Fietsbarometer (Storme et al., 2022), worden de resultaten van beide onderzoeken vergeleken met elkaar. Dit gebeurt aan de hand van een Quadrat Analysis (QA) met een Chi-square Goodness-of-Fit test ( $\chi^2$ ) (Thomas, 1977) zoals ook gedaan werd door Ramírez Aranda et al. (2021). Hiervoor stellen we onze nulhypothese  $H_0$  dat de ruimtelijke verdeling van de zones die als gevaarlijk zijn aangeduid met behulp van de LIM methode gelijk zijn met de ruimtelijke verdeling van de wegsegmenten die als gevaarlijk zijn beoordeeld met behulp van de Fietsbarometer. De alternatieve hypothese  $H_1$  is dat de ruimtelijke verdeling van de twee zones verschillend zijn.

QA is een techniek ontwikkeld in de ecologie waarbij er een grid wordt gelegd over het studiegebied, en per gridcel worden het aantal waarnemingen geteld (Rogerson, 2015). Het is dan wel belangrijk dat er een goede grootte wordt gebruikt voor de gridcellen zodat er een maximum aan informatie behouden blijft. Als de cellen te klein zijn, zullen er veel lege cellen zijn en zullen veel

clusters gemist worden. Als de cellen te groot zijn, kan men patronen missen die zich binnenin één cel afspeelt. Om een celgrootte te vinden die ons de meeste informatie geeft, gebruiken we Griffith et al.'s (1990) formule voor de celgrootte.

$$\text{Quadrat groote} = \frac{2 * A}{n} \quad \Rightarrow \quad \text{Lengte zijkant gridcell} = \sqrt{\frac{2 * A}{n}}$$

waar A de oppervlakte is van het studiegebied, en n het totale aantal observaties van de dataset met de minste observaties.

Per gridcel wordt geteld hoeveel aangeduiden gevarenzones van de LIM methode erin vallen of mee overlappen. Dit is de geobserveerde verdeling O. Voor de Fietsbarometer wordt per gridcel geteld hoeveel beoordelingen zijn die een wegsegment beordelen die door die gridcel loopt. Dit is de verwachte verdeling E.  $\chi^2$  gaat ervan uit dat we met geobserveerde telgegevens werken en deze zogezegd vergelijk met een verwachte verdeling waarbij er evenveel observaties zijn als in de geobserveerde verdeling. Het aantal observaties in beide datasets verschillen echter. Daarom worden de geobserveerde en verwachte verdeling herschaald naar het aantal observaties van de verdeling met het minst aantal observaties. Dan kan de  $\chi^2$  Goodness-of-Fit berekend worden met:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

De p-waarde kan dan gehaald worden uit een  $\chi^2$ -verdeling tabel. Bij een significante p-waarde wordt  $H_0$  verworpen.

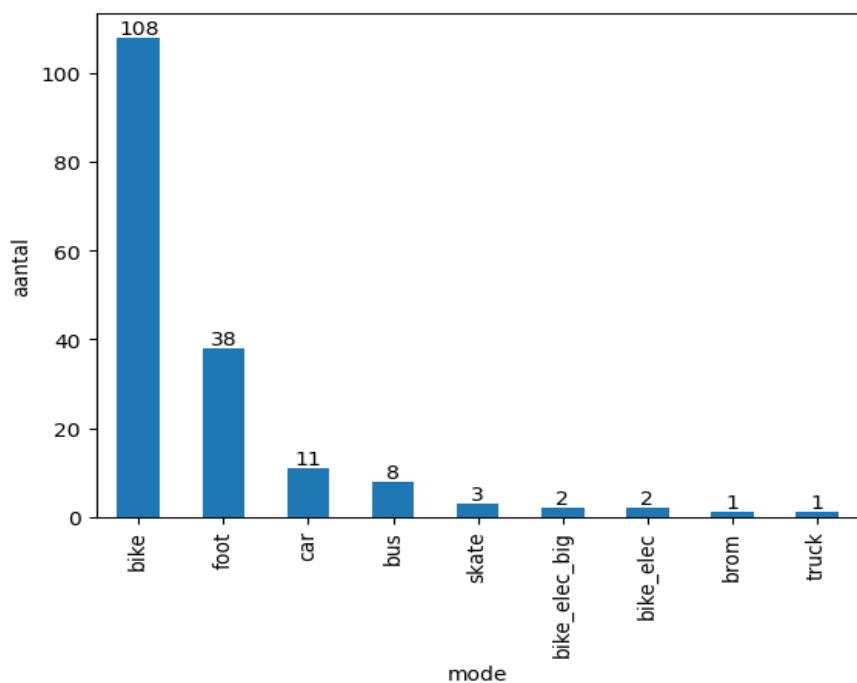
Uiteindelijk worden ook de resultaten van de Fietsbarometer en LIM visueel vergeleken met elkaar. Hieruit besluiten de onderzoekers welke van de twee methoden subjectief beter is. Dit is beter op vlak van compleetheid, details,

accuraatheid, en inzichten die kunnen verworven worden eruit. De Fietsbarometer heeft als voordeel dat straatsegmenten apart worden beoordeeld, en dat uitgebreid wordt gevraagd naar verschillende parameters die de perceptieve veiligheid kunnen beïnvloeden. Het heeft dan wel als nadeel dat de respondent vasthangt aan de wegsegmenten die ze kunnen beoordelen, en de redenen die gegeven worden door de onderzoeker. LIM geeft de vrijheid aan de respondent om aan te duiden en aan te geven wat deze wilt en kan zo meer specifieke details opvragen. Dit kan echter zeer chaotisch ingevuld worden door de respondenten, en de verkregen antwoorden zijn niet noodzakelijk de antwoorden op zaken die de onderzoeker zou willen onderzoeken.

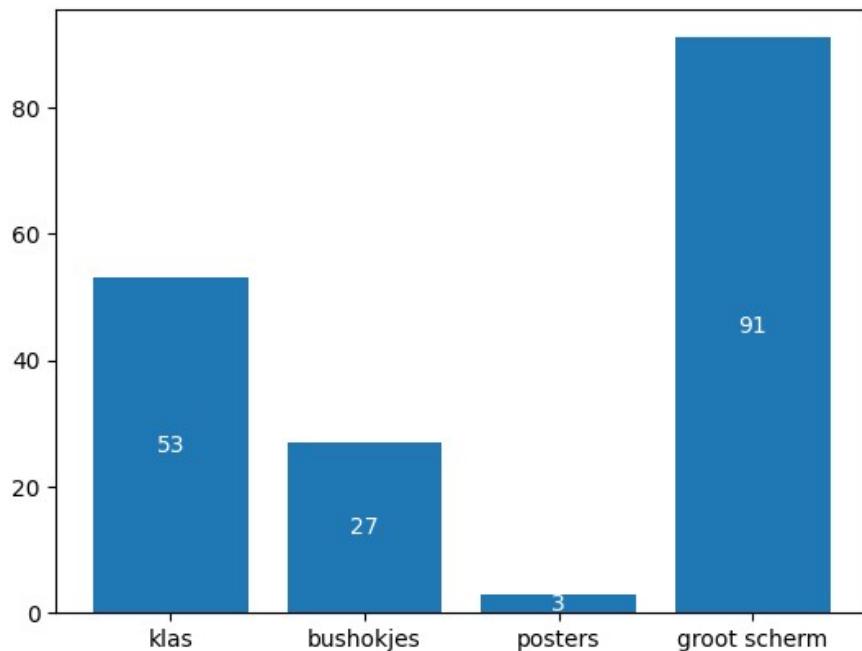
### **3 RESULTATEN**

#### **3.1 Meta statistieken**

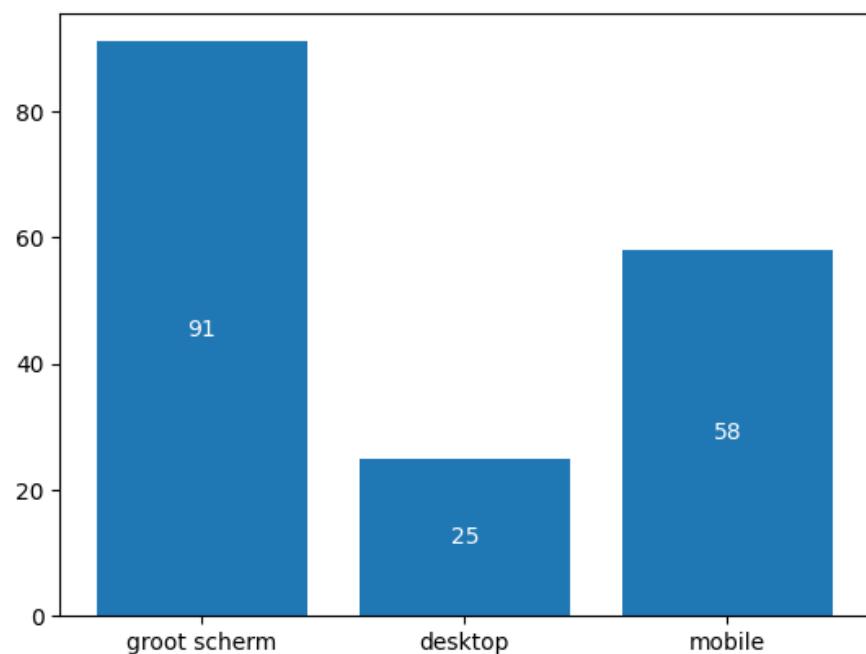
Er zijn 174 bruikbare inzendingen op de vragenlijst. Er waren een aantal inzendingen die als grap gemaakt werden, en deze zijn er manueel uit verwijderd. Van de bruikbare inzendingen beoordelen de meeste (108) de fiets als vervoersmiddel, gevolg door te voet (38), en de auto (11) en met de bus (8) (Figuur 4). Van alle andere vervoersmodi zijn er telkens maar enkele inzendingen. Figuur 5 toont hoe dat mensen bij de vragenlijst zijn gekomen. Het groot touchscreen aan de inkom van de S8 was het meest succesvol. Deze trekt snel de aandacht en zet de mensen aan om dit in te vullen. Het was ook succesvol om aan de studenten in de klas, en aan andere mensen in de vakgroep, persoonlijk te vragen om dit in te vullen. Flyers en posters hangen aan bushokjes en bij handelszaken was niet zo succesvol.



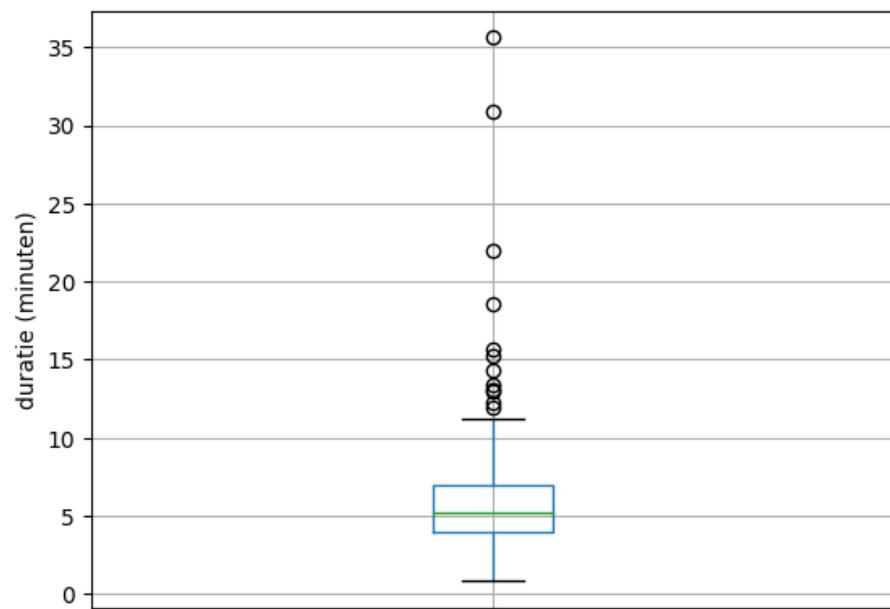
**Figuur 4: Verdeling vervoersmodi**



**Figuur 5: Oorsprong reclame**



**Figuur 6: Mobile vs Desktop**



**Figuur 7: Tijdsduur invullen van de vragenlijst**

Aangezien vele mensen via de vragenlijst zijn gekomen via het groot touchscreen, hebben dus ook veel mensen de vragenlijst ingevuld met het

groot touchscreen. Dit groot touchscreen is moeilijk te categoriseren onder desktop of mobile, dus is dit een aparte categorie in Figuur 6. Tussen desktop en mobile hebben de meeste mensen de vragenlijst ingevuld met hun mobiel toestel, waarvoor het ontworpen is. Het is raar dat nog een redelijk aantal respondenten de vragenlijst op een desktop PC of laptop hebben ingevuld, aangezien de reclame voornamelijk gericht was op situaties waar je je mobiele telefoon gebruikt.

Figuur 7 toont hoe lang respondenten erover deden om de vragenlijst in te vullen. Gemiddeld deden ze er vijf minuten over, wat meer is dan de verwachte drie minuten. Het lagere kwantiel is echter de verwachte drie minuten. Waarschijnlijk hadden mensen meer tijd nodig om te denken over de open vragen, of uit te zoeken hoe het tekenen op de kaart werkt. Er zijn wel een heel deel uitschieters die er vijftien tot dertig minuten over deden. Eén respondent deed er zelfs vier en een half uur over, deze wordt niet getoond op het boxplot aangezien de rest dan niet meer leesbaar is. De uitschieters zijn waarschijnlijk van het grote touchscreen dat mensen zijn beginnen invullen, maar er dan mee stopten en dat iemand anders het later verder invulde. Alhoewel het touchscreen de vragenlijst herstart na drie minuten aan inactiviteit.

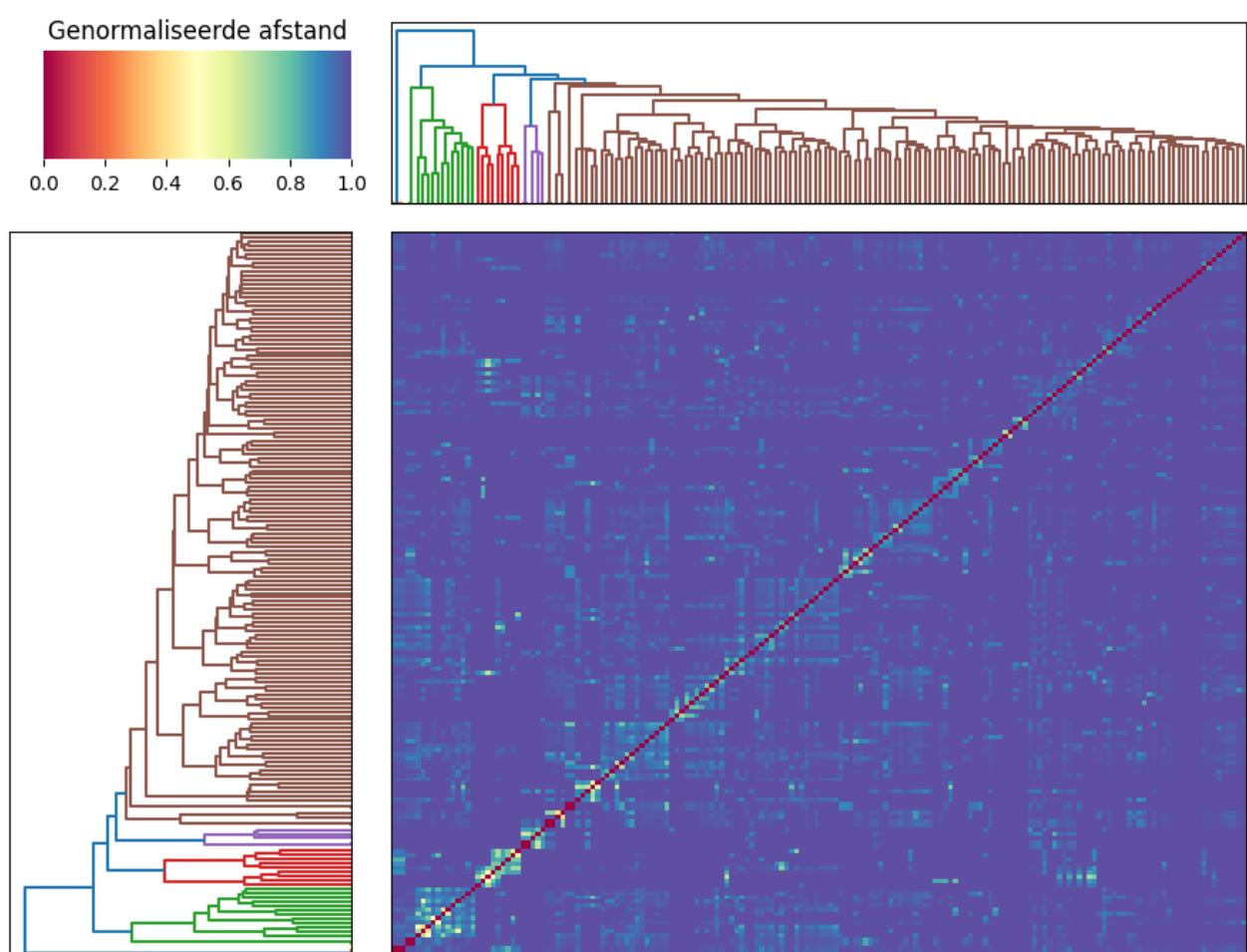
## 3.2 Topics uit open vragen

### 3.2.1 Vraag 4: wat is er gevaarlijk aan?

Figuur 8 toont een wordcloud van de antwoorden op Vraag 4. Woorden over de verschillende vervoersmodi zoals “fietser”, “auto”, en “voetganger” zijn zeer prominent aanwezig. Woorden die specifiek zijn aan het studiegebied zelf zijn ook te herkennen, zoals “tunnel” en “kruispunt”. De overige veelvoorkomende woorden kunnen al een indicatie geven van de verschillende topics die we zouden kunnen vinden, zoals “snel/snelheid”, “voorrang”, “zien/zicht/kijken”, “verkeer”, “oversteken”.

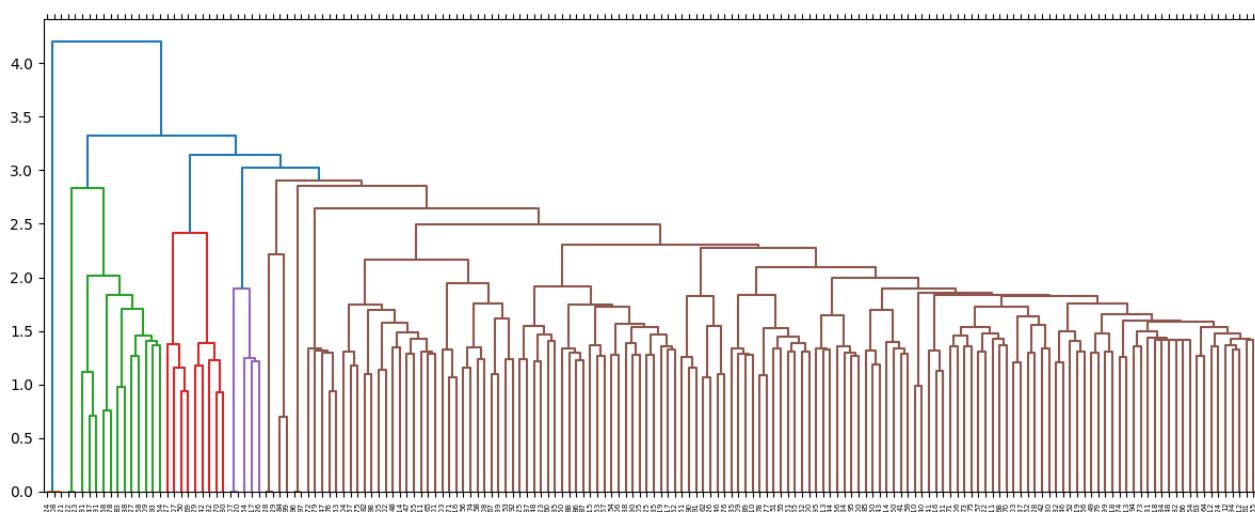


Figuur 8: Wordcloud Vraag 4

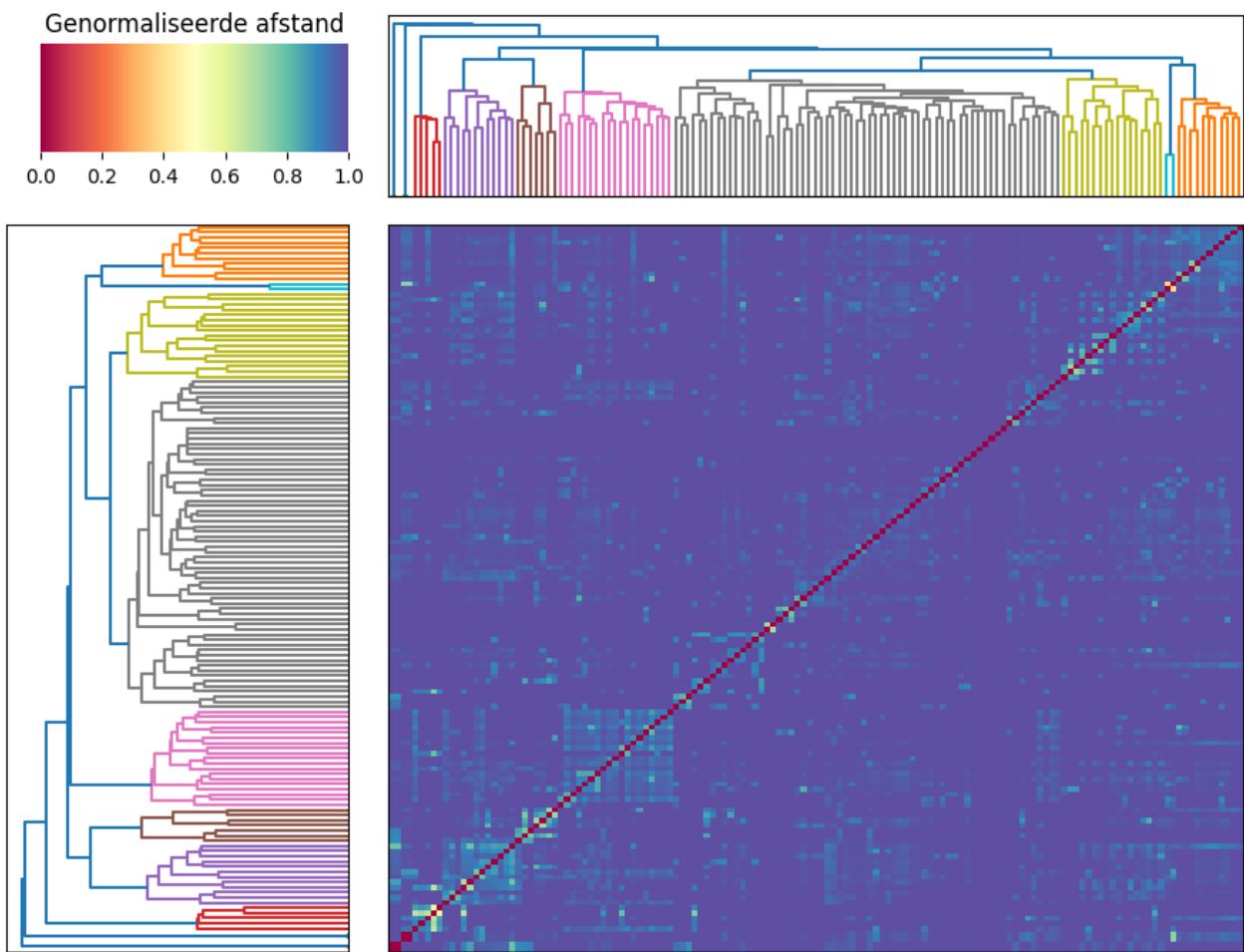


Figuur 9: Heatmap en dendrogram Vraag 4, eerste iteratie

Na het berekenen van de afstand met de cosinus dissimilariteit en het bepalen van de hiërarchische koppelingen met Ward's methode, is in Figuur 9 de complexe clusterstructuur zichtbaar. Figuur 10 focust nog eens op enkel het dendrogram. In het begin worden er al een aantal clusters onderscheiden met een duidelijke threshold. De eerste groep zijn drie identieke antwoorden die praten over dat het een punt is waar veel verkeerstromen samenkommen. De tweede cluster, de groene, gaan allemaal over de snelheid van het verkeer. De derde groep, de rode, gaan over het feit dat er veel verkeer is. De vierde groep, de paarse, gaan over het feit dat ze daar moeten oversteken. Als we echter de vijfde groep, de bruine, met "andere" verder willen opdelen door de threshold te verlagen, zal de groene cluster en rode cluster opgesplitst worden welke al goed gedefinieerde clusters waren. Daarom wordt voor enkel de laatste groep de clustering opnieuw gedaan met een lagere, fijnere threshold.



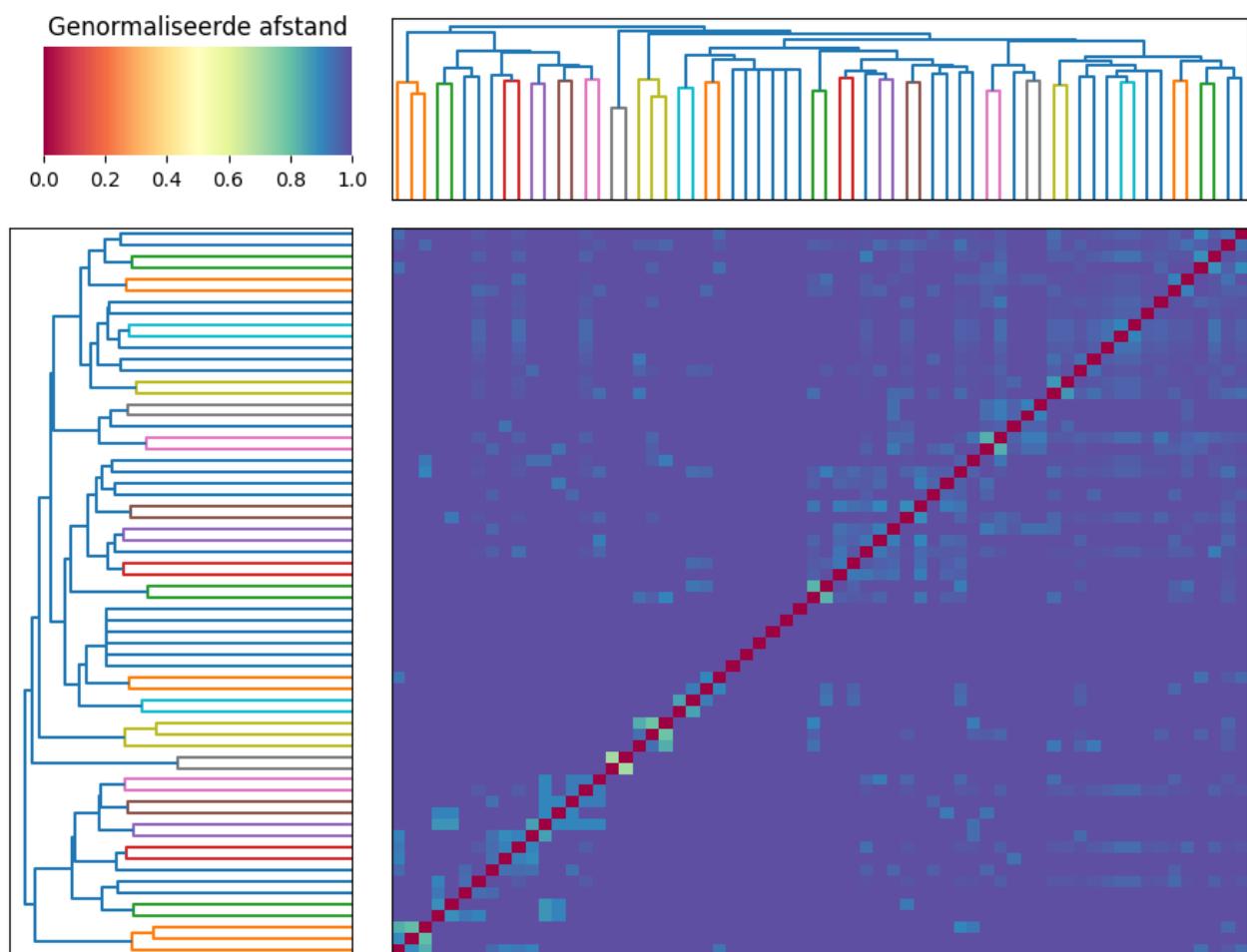
**Figuur 10: Dendrogram Vraag 4, eerste iteratie**



**Figuur 11: Heatmap en dendrogram Vraag 4, tweede iteratie**

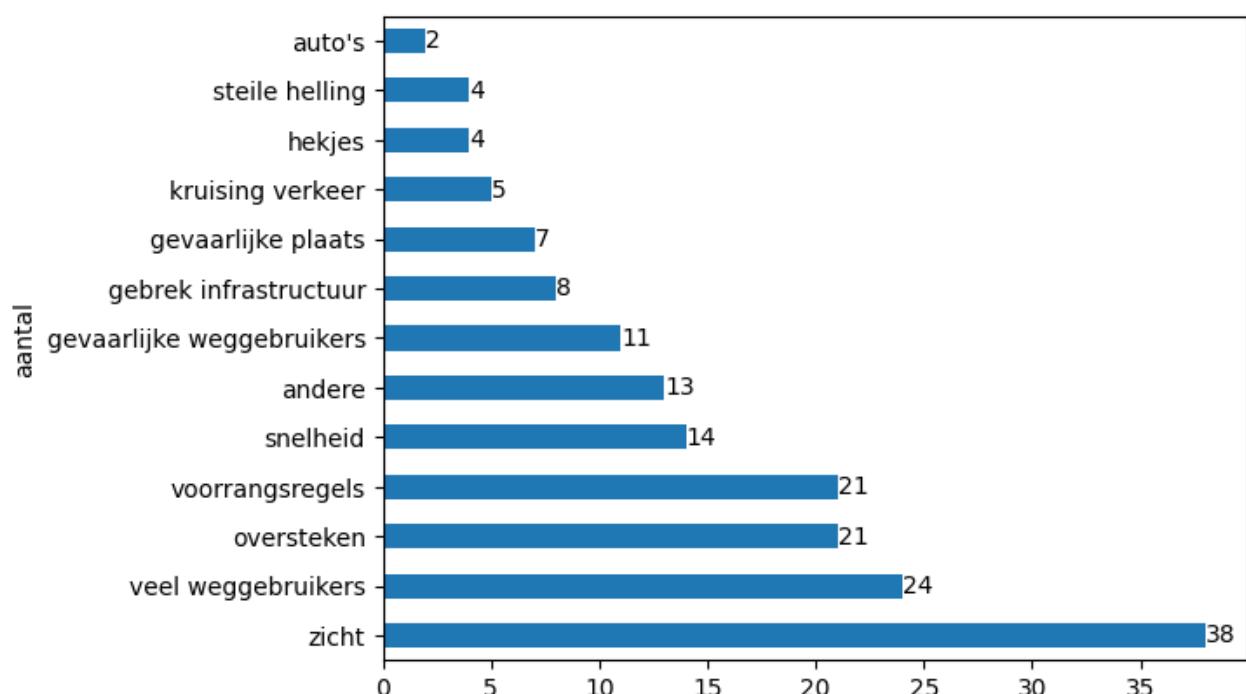
Bij de tweede iteratie toont de heatmap in Figuur 11 weer duidelijke clusters zoals aangegeven wordt door het dendrogram, alsook één grote groep in het midden waar er op de heatmap nog wat verborgen clusters zitten, dus een derde iteratie zal nog nodig zijn. In de tweede iteratie komt het topic “zicht” regelmatig voor in clusters die niet altijd dicht bij elkaar liggen. “Zicht” komt voor in clusters 1, 4, 5 en 8. Dit komt omdat het vaak op verschillende manieren verwoord wordt. Bijvoorbeeld in cluster 1 spreekt men over “niet zien”, en in cluster “8” spreekt men over “slecht zicht”. Alhoewel dit over hetzelfde praat, wordt het op een totaal andere manier gezegd.

Bij de derde iteratie is het al een stuk moeilijker om duidelijke grote clusters te onderscheiden, zoals te zien is op Figuur 12. We hebben gekozen om dit nog op te splitsen in 25 clusters. Vele van deze clusters praatten over hetzelfde, maar het is moeilijker om een ideale threshold te vinden dan om clusters later samen te voegen. Ook waren er een paar clusters waarvan de antwoorden niets met elkaar te maken hadden, ook al zijn er maar twee antwoorden per cluster. Ook was het soms moeilijk om de antwoorden apart te categoriseren bij een al bestaande topic. Deze clusters hebben dan geen topic gekregen. Het kan wel interessant zijn om deze antwoorden nog eens apart te bekijken als uitschieter, maar dit is hier niet gedaan.



**Figuur 12: Heatmap en dendrogram Vraag 4, derde iteratie**

Uiteindelijk werden alle clusters die over hetzelfde topic spreken samengevoegd, en bleven er nog 12 topics over. De gevonden topics zijn: zicht, snelheid, oversteken, veel weggebruikers, auto's, gevaarlijke weggebruikers, gevaarlijke plaats, steile helling, voorrangsregels, kruising verkeer, gebrek infrastructuur, en hekjes. Figuur 13 toont de verdeling van het aantal antwoorden per topic. Opvallend is dat er meer mensen praten over het gedrag van andere weggebruikers en over algemene zaken, dan over infrastructuur en dingen eigen aan de locatie.

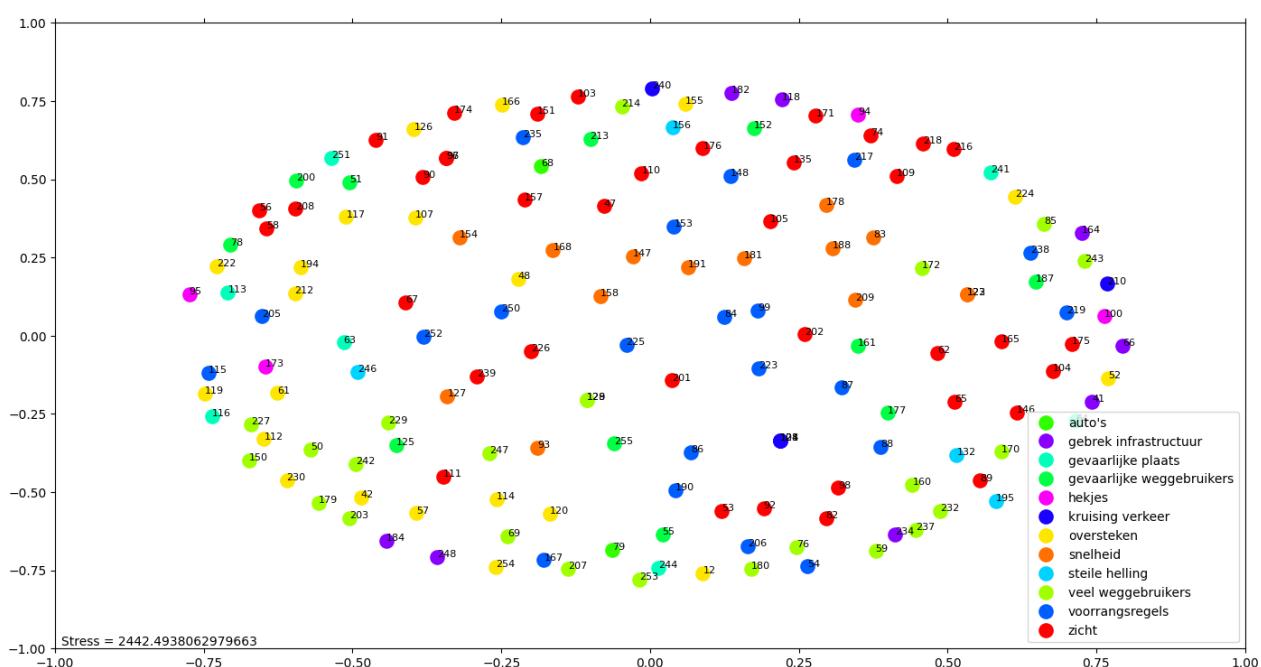


**Figuur 13: Aantal antwoorden per topic van Vraag 4**

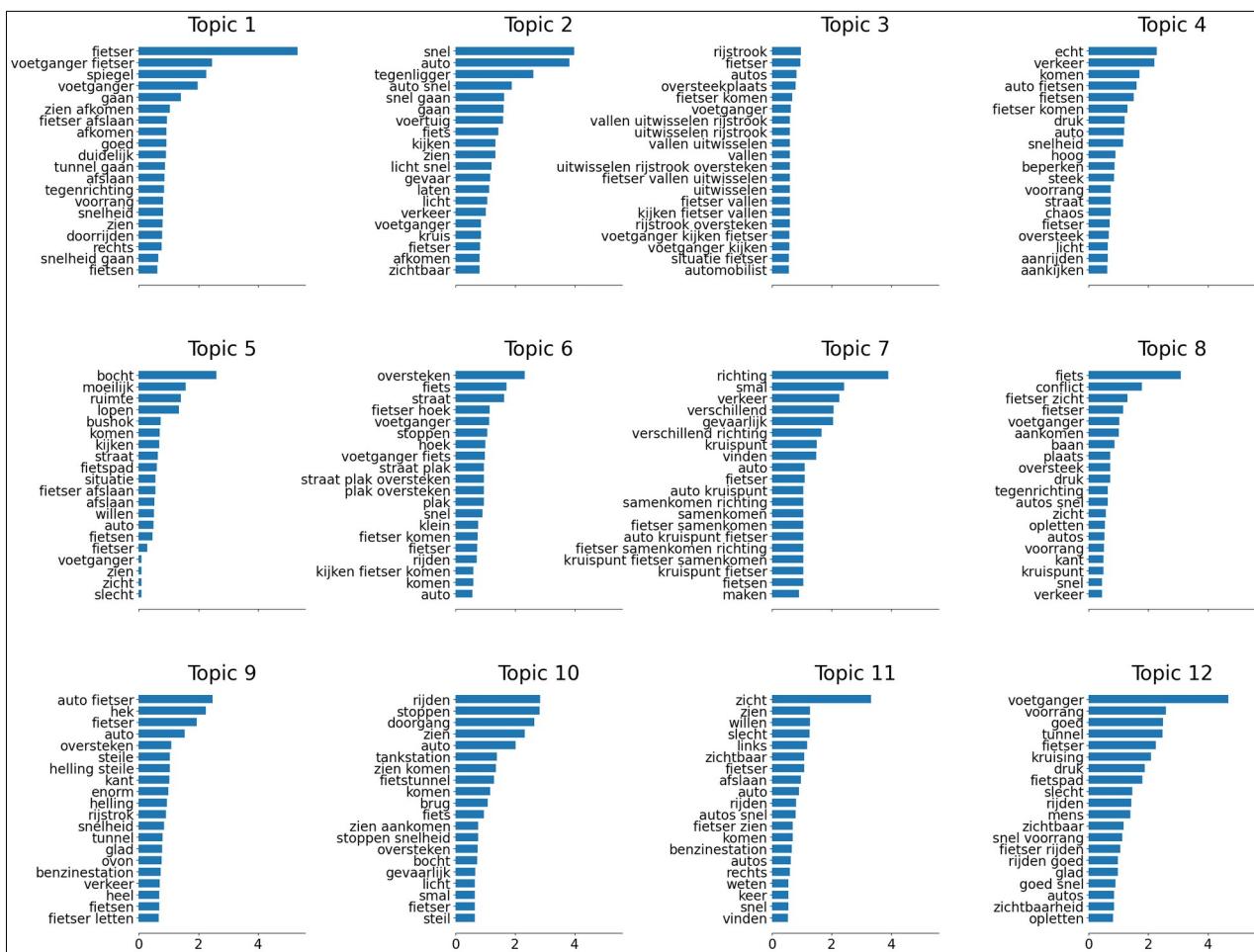
### 3.2.1.1 Evaluatie van de topics

Om de verkregen topics te bekijken we nog het MDS-plot. Het MDS-model heeft een uiteindelijk stress factor van 2442.49, wat betekent dat het een zeer slechte fit is. Dit is wel te verwachten aangezien we een zeer complexe, multidimensionale vectorruimte ( $n=172$ ) met een complexe afstandsmaat, de cosinus dissimilariteit, proberen te passen in een 2-dimensionele ruimte

volgens euclidische afstanden. Ook bevat de dataset veel ruis, wat ervoor kan zorgen dat de MDS kortsluiting maakt (Peterfreund & Gavish, 2021). MDS scherpstellen op data met veel ruis is nog een actief onderzoeksgebied. Vanwege de hoge stress en de ruis in de data, mogen we de onderlinge afstanden in het MDS plot (Figuur 14) niet al te serieus nemen. Toch kunnen we de verkregen topic clusters een beetje onderscheiden als we ze inkleuren, al vormen ze geen duidelijke clusters in het MDS-plot. Het topic “snelheid” (oranje) zit bijvoorbeeld ongeveer allemaal bij elkaar. Het topic “zicht” (rood) bestaat uit allemaal verschillende clusters, want deze hebben we ook gevormd uit allemaal verschillende clusters. De topics “gebrek infrastructuur” en “hekjes” zitten beiden verspreid. Hier zit er waarschijnlijk weinig coherentie tussen in de manier hoe dat de antwoorden zijn geformuleerd.



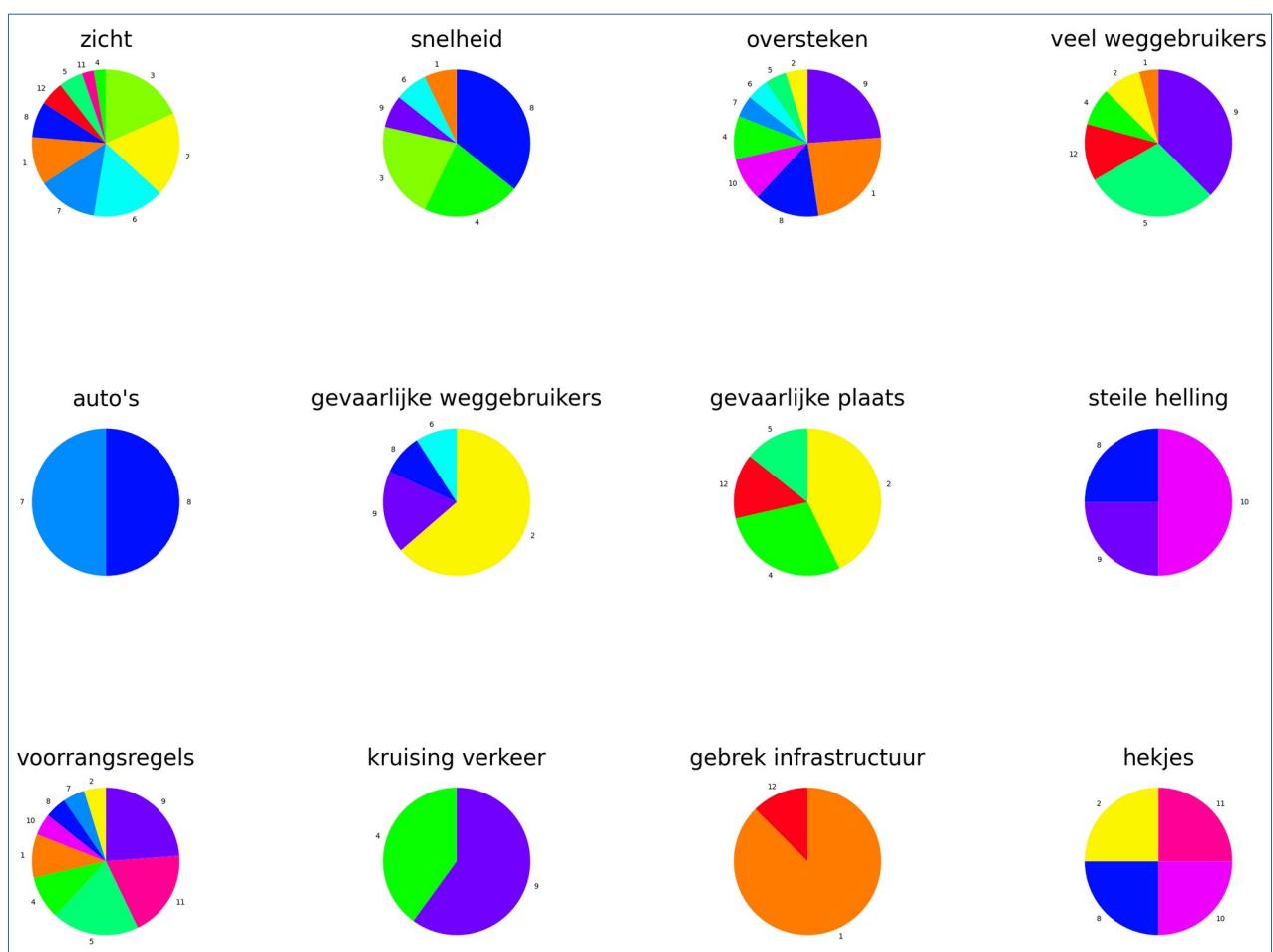
**Figuur 14: MDS plot van de topics van Vraag 4**



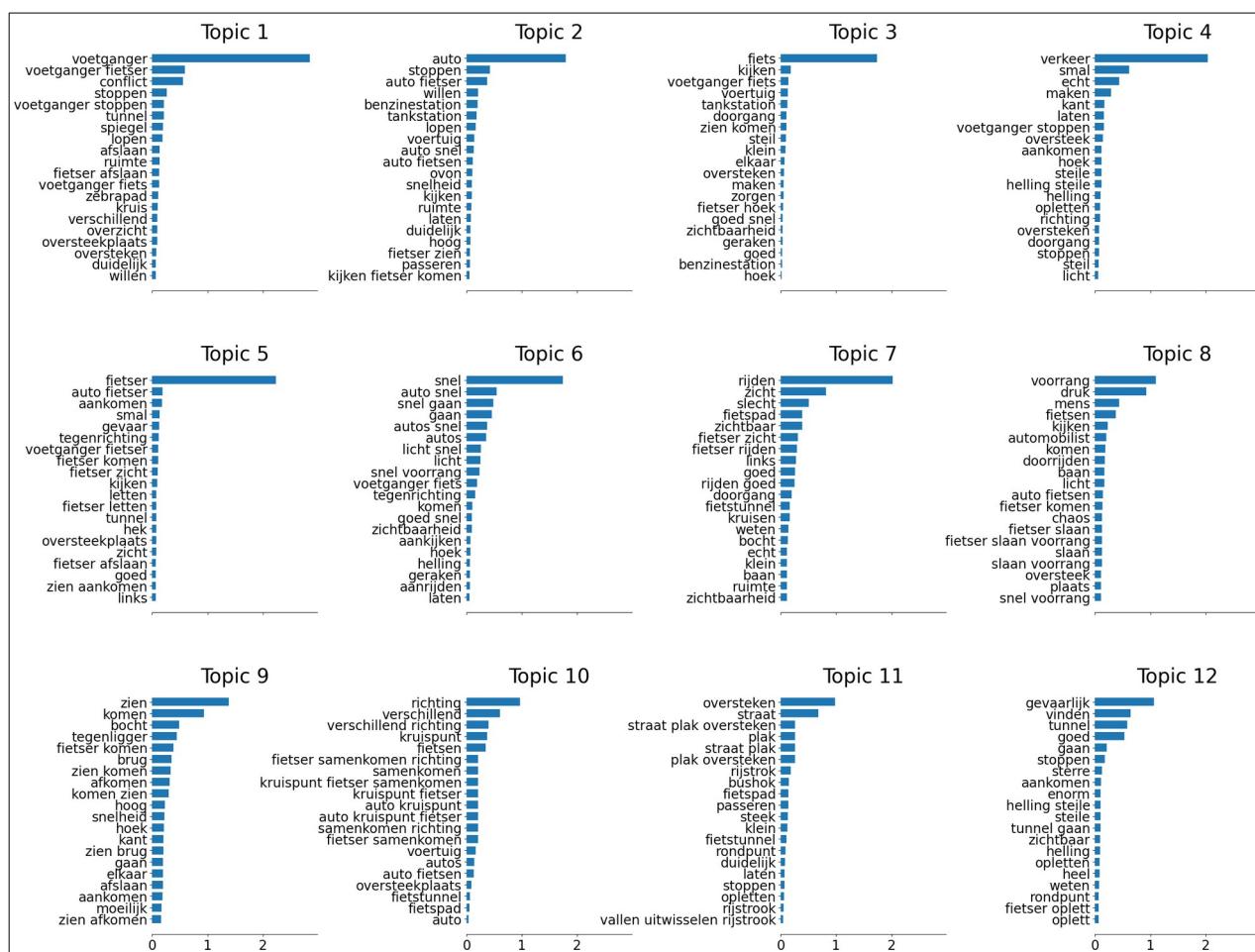
**Figuur 15: LDA topics voor Vraag 4**

Figuur 15 toont de LDA-topics voor Vraag 4 als we ervan uitgaan dat er 12 topics zijn zoals we hebben gevonden in de Hiërarchische clustering. Topic 1 kan mikken op de interactie tussen fietser en voetganger. Topic 2 zou kunnen gaan over de snelheid van de auto's. Zowel topic 4 als topic 11 gaan over het zicht. Topic 6 gaat over oversteken en topic 12 zou gaan over de voorrang van voetgangers. De andere topics zijn al wat moeilijker om te interpreteren. Een vergelijking tussen de LDA-topics en de HC-topics wordt gemaakt in Figuur 16. Veelvoorkomende HC-topics zoals "zicht", "snelheid", "oversteken", "veel weggebruikers", en "voorrangsregels" bestaan uit meerdere tot bijna alle LDA-topics. Als we wat kijken naar de algemene kleurenopmaak van de pie charts, zou je overeenkomsten tussen topics

kunnen zien. "Gevaarlijke plaats" lijkt zo gelinkt te zijn aan "kruising verkeer" en "steile helling". "Gevaarlijke weggebruikers" lijkt ook wat verbanden te hebben met "snelheid" en "voorrangsregels". Deze vergelijking is echter puur subjectief, maar men zou kunnen denken dat er logische verbanden achter zitten. Een plaats kan gevaarlijk zijn door de steile helling of het verkeer dat er kruist en weggebruikers kunnen gevaarlijk zijn door hun snelheid en omdat ze de voorrangsregels niet goed opvolgen.

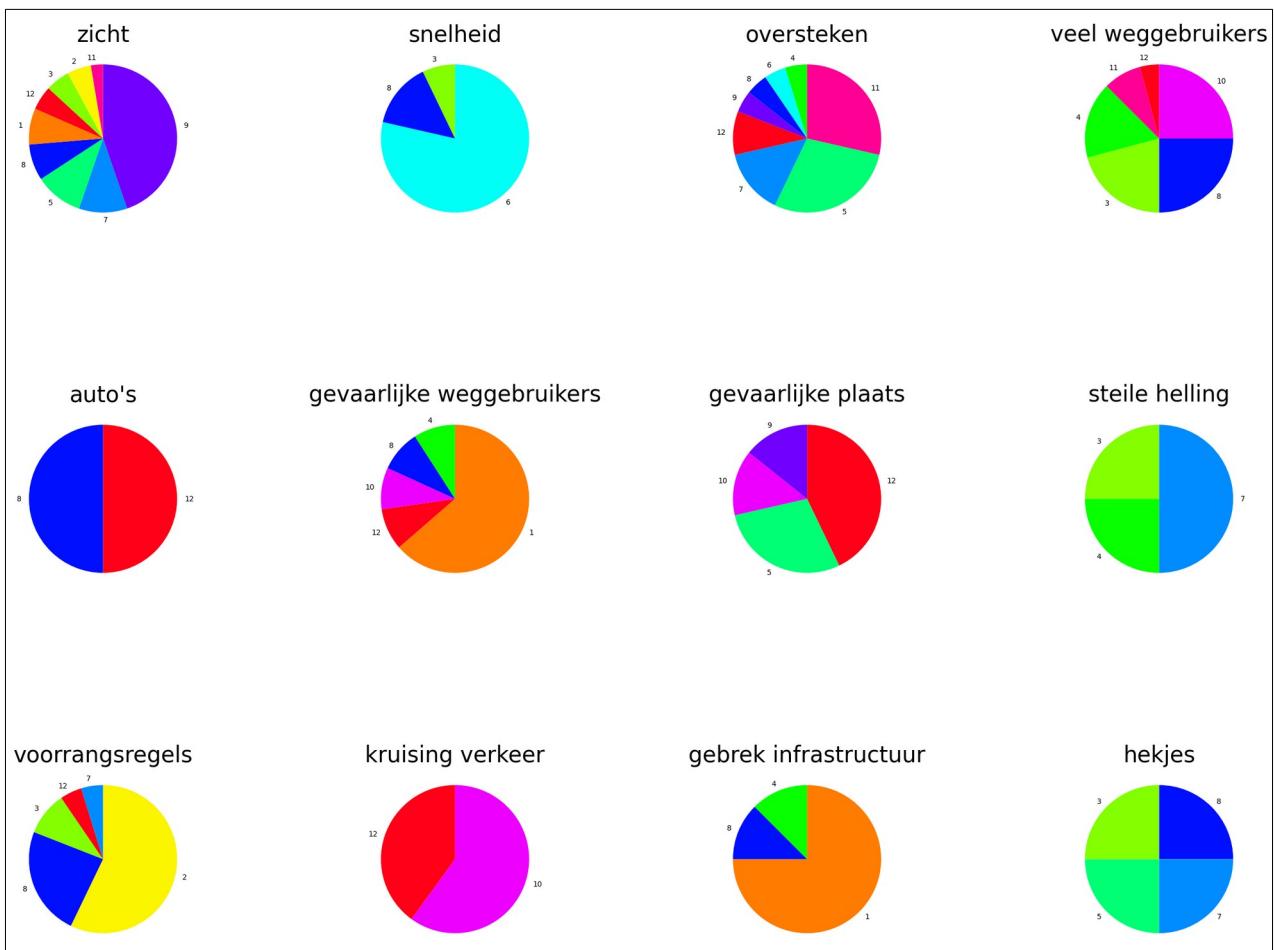


**Figuur 16: Vergelijking LDA topics per HC topic voor Vraag 4**



**Figuur 17: NMF topics voor Vraag 4**

Figuur 17 toont de NMF-topics voor Vraag 4. Topic 1 zou kunnen gaan over voetgangers, topic 2 over auto's en topics 3 en 5 over fietsers. Topic 6 zou kunnen gaan over snelheid, topic 8 over voorrang en drukte, en topics 7 en 9 over zicht. Deze topics lijken wel al beter gedefinieerd dan de LDA-topics, en dat is ook te zien in de vergelijking met de HC-topics in Figuur 18. Hier is er meer één NMF-topic per HC-topic dat domineert, en bestaan de HC-topics over het algemeen over minder verschillende NMF-topics. Je zou, als je naar de overeenkomstige kleuren kijkt, kunnen zeggen dat "gevaarlijke plaats" en "oversteken" verbanden hebben. Een plaats waar je moet oversteken kan je namelijk ook als gevaarlijk beschouwen, maar dit is opnieuw subjectief gekeken.



**Figuur 18: Vergelijking NMF topics met HC topics voor Vraag 4**

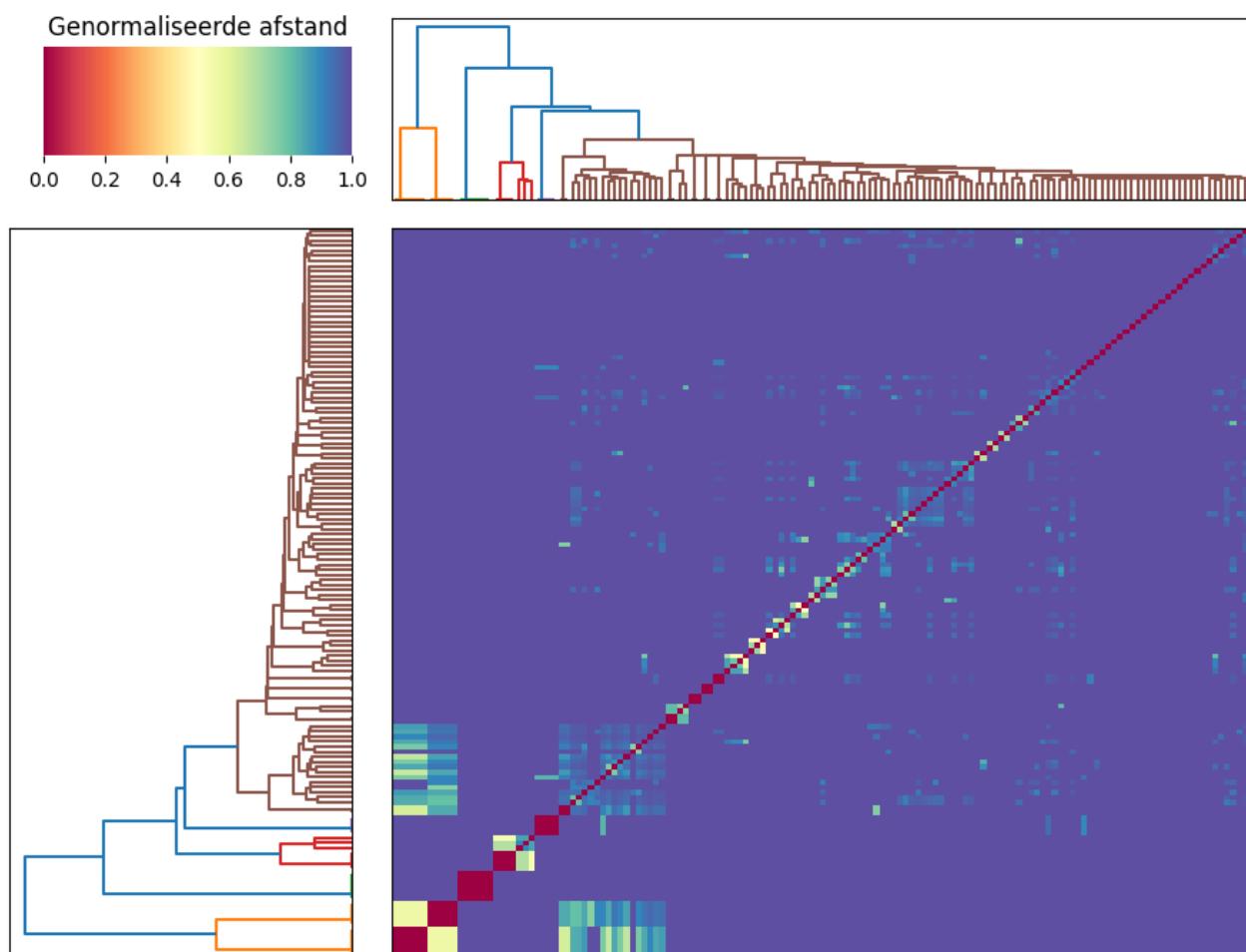
### 3.2.2 Vraag 5: hoe zou u het verbeteren

Figuur 19 toont de wordcloud voor de antwoorden op vraag 5. Weer komen de vervoersmodi, “fiets”, “auto”, “voetganger”, er prominent in voor. Het woord dat het meeste voorkomt is “spiegel”. We kunnen dus al verwachten dat veel mensen praten over het plaatsen van spiegels, en dat “spiegel” dus één van de topics zal zijn. Andere interessante woorden in de wordcloud zijn “tankstation”, “fietspad”, “tunnel”, “verkeerslichten”, “bord” en “licht”.

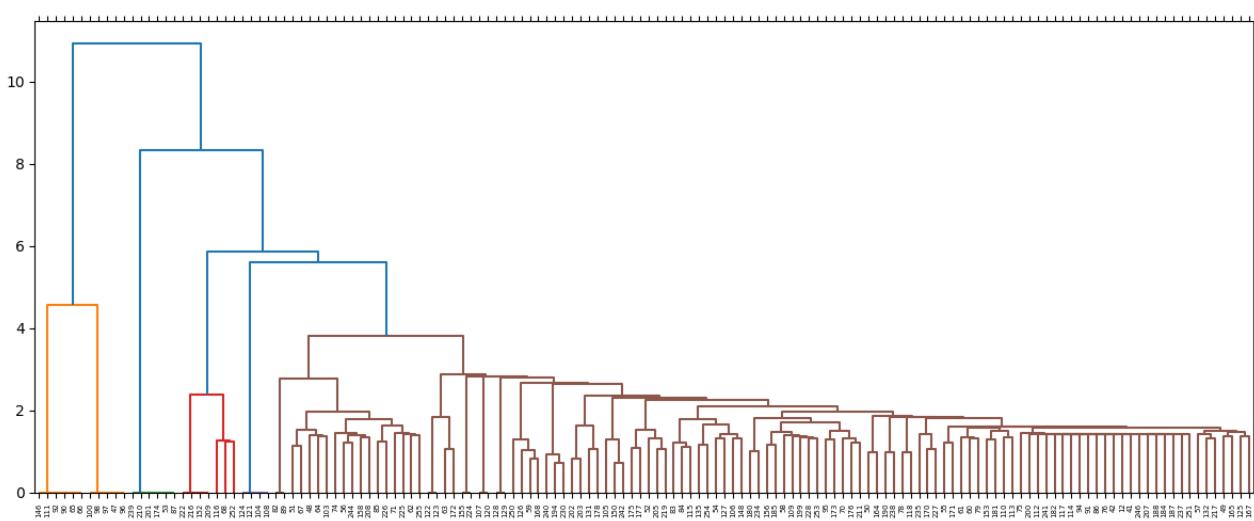


**Figuur 19: Wordcloud Vraag 5**

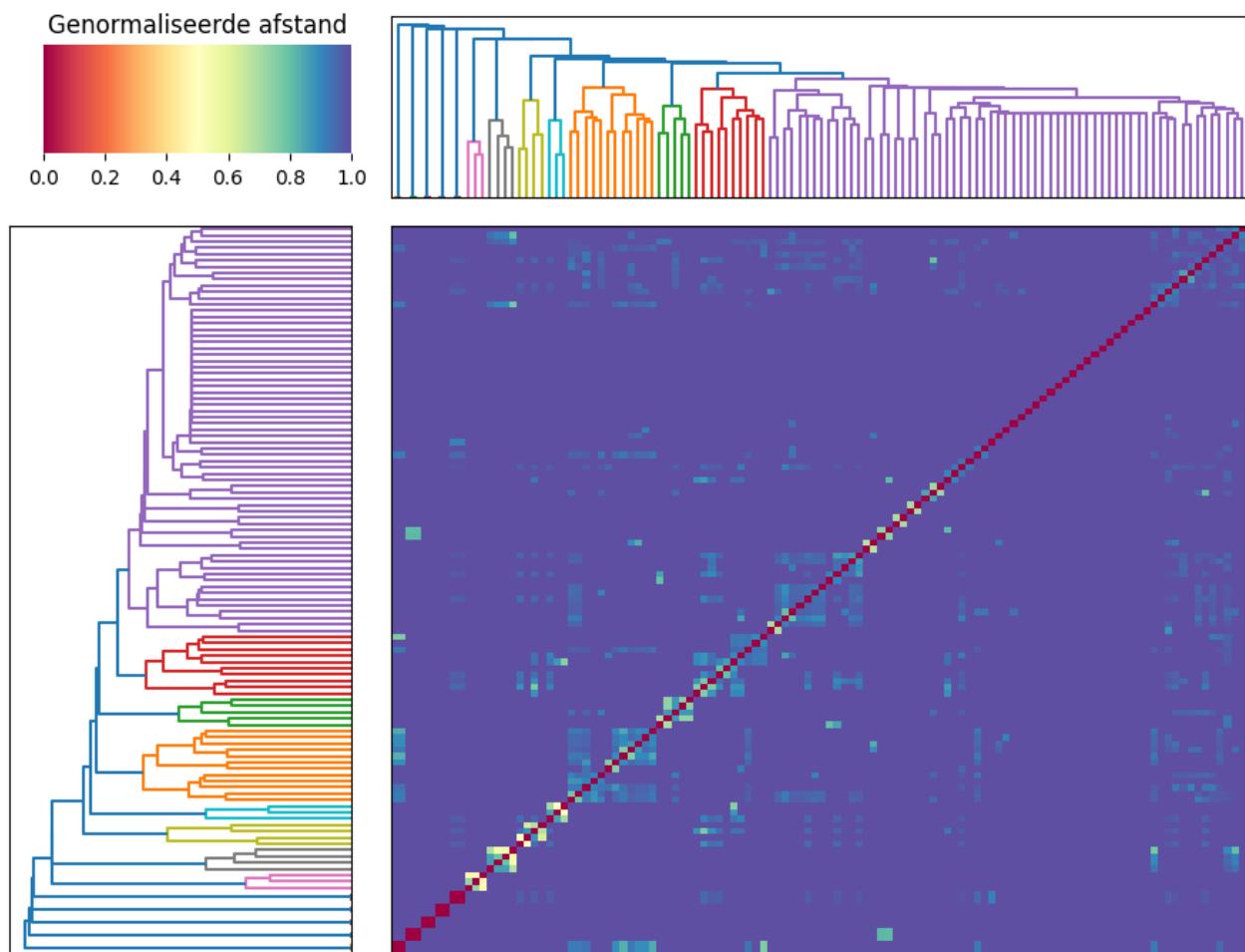
De heatmap in Figuur 20 toont zeer duidelijke clusters, gevuld door een paar minder duidelijke clusters. We kunnen al vijf clusters afsplitsen voor dat we de eerste cluster ongewild in twee splitsen. De eerste twee clusters gaan allebei over het plaatsen van spiegels. Het enige verschil is dat de tweede cluster in het Engels staat. Normaalgezien zijn alle antwoorden vertaald naar het Nederlands in de data opkuis stap, maar dit is blijkbaar niet perfect gebeurd. De derde cluster zijn allemaal mensen die zeggen dat ze geen idee hebben hoe we de verkeerssituatie kunnen verbeteren. Alhoewel deze vraag optioneel was, hebben mensen dit dus toch zo ingevuld waarschijnlijk omdat de vorige vraag een verplichte open vraag was. De vierde cluster spreekt over het plaatsen van verkeerslichten en de vijfde cluster is nog een cluster “andere” die verder opgesplitst wordt.



**Figuur 20: Heatmap en dendrogram van Vraag 5, eerste iteratie**

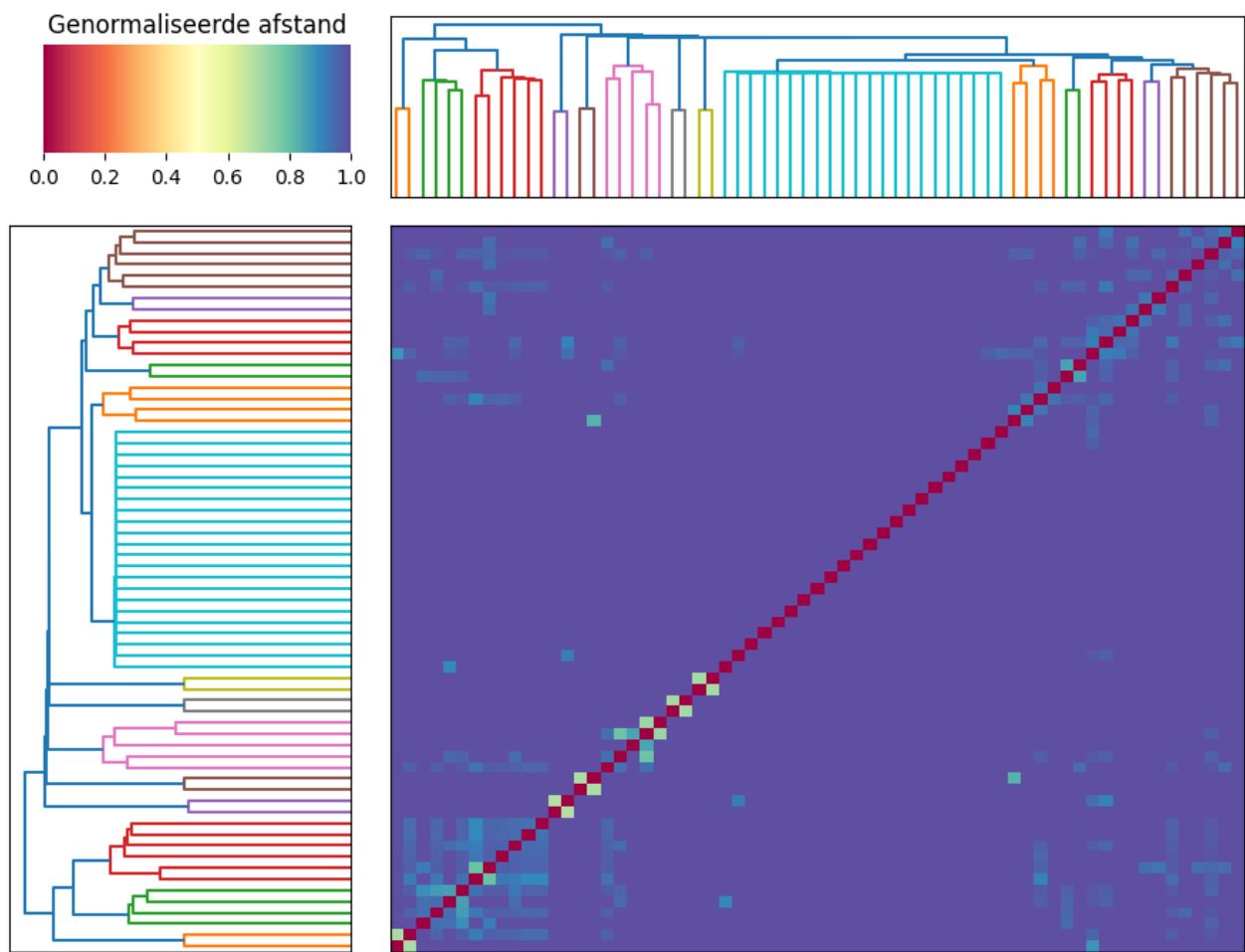


**Figuur 21: Dendrogram van Vraag 5, eerste iteratie**



**Figuur 22: Heatmap en dendrogram Vraag 5, tweede iteratie**

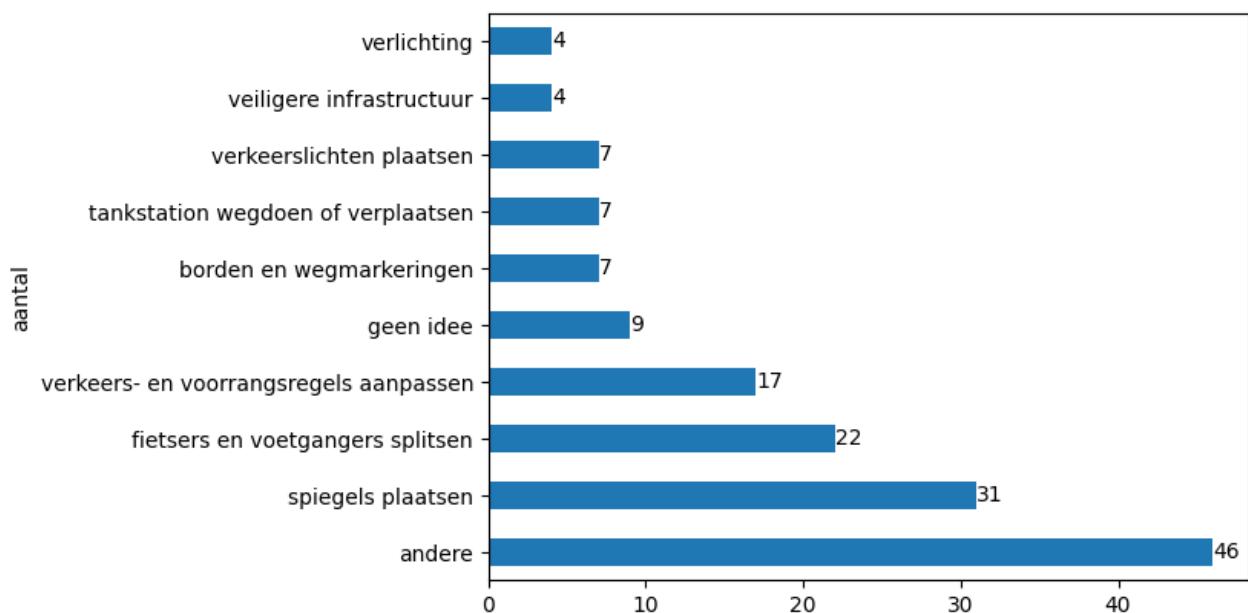
Bij de tweede iteratie in Figuur 22 zijn er nog een paar licht duidelijke clusters te onderscheiden. We hebben deze iteratie gesplitst in 13 clusters. Weer al komen te topics van “spiegels plaatsen” en “verkeerslichten plaatsen” voor, juist een beetje anders verwoord, bijvoorbeeld met “zetten” in plaats van “plaatsen”. Vaak komt ook het topic aan over het aanpassen van de verkeersregels in de vorm van allerlei voorbeelden die toch samen geclusterd worden. De cluster “andere” begint nu ook voornamelijk veel kleine interne clusters te tonen. Dit betekent dat we nu enkel nog de ruis moeten clusteren.



**Figuur 23: Heatmap en dendrogram Vraag 5, derde iteratie**

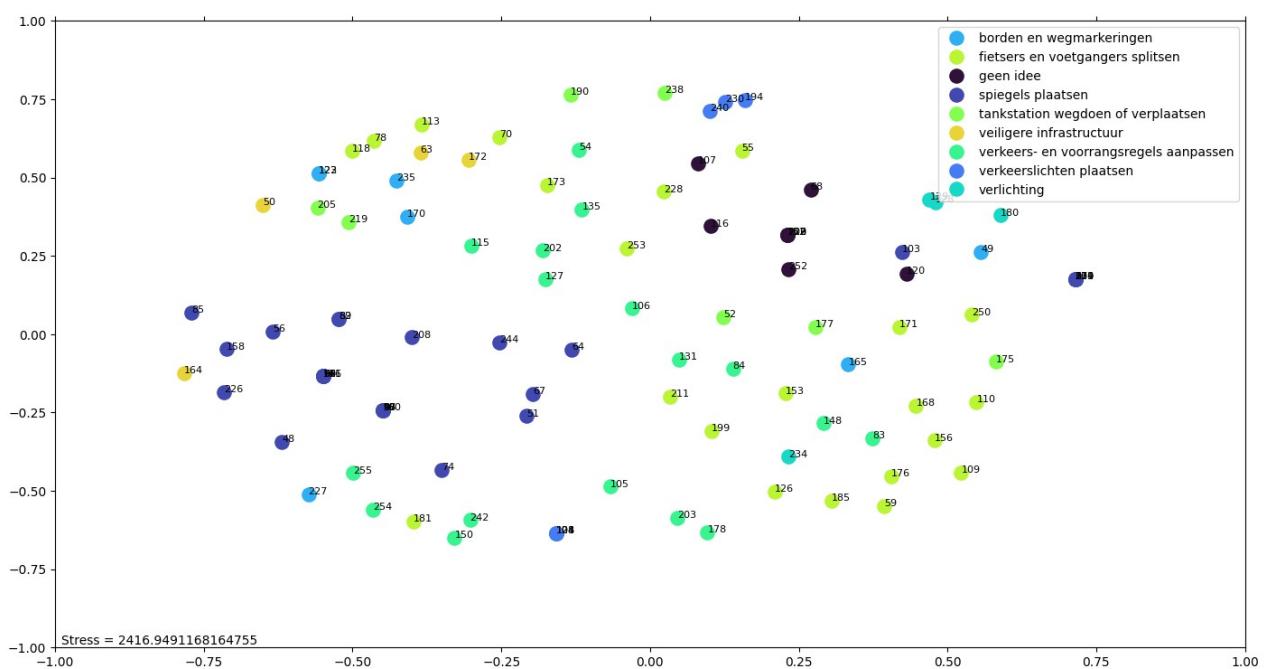
In de derde iteratie (Figuur 23) zijn de clusters minder sterk gedefinieerd. We identificeren nog 14 clusters. Aan de heatmap is te zien dat de grote lichtblauwe cluster enkel ruis is en geen onderliggende verbanden heeft. Ook is er in de laatste en derde laatste cluster geen duidelijk overliggend topic. Men zou manueel nog het topic kunnen definiëren per antwoord waar er al een duidelijk geïdentificeerd topic is, maar dit is hier nu niet gedaan. De uiteindelijk gevonden topics met de hiërarchische clustering zijn: spiegels plaatsen, verkeerslichten plaatsen, borden en wegmarkeringen, verlichting, verkeers- en voorrangsregels aanpassen, tankstation verplaatsen of wegdoen, aparte plaats voor fietsers en voetgangers, veiligere infrastructuur, en geen idee. Het grootste aantal topics is “andere” (Figuur 24), gevolgd door

“spiegels”. Het kan dus nog zeer interessant zijn voor de domeinspecialist om in detail naar de antwoorden van topic “andere” te kijken.



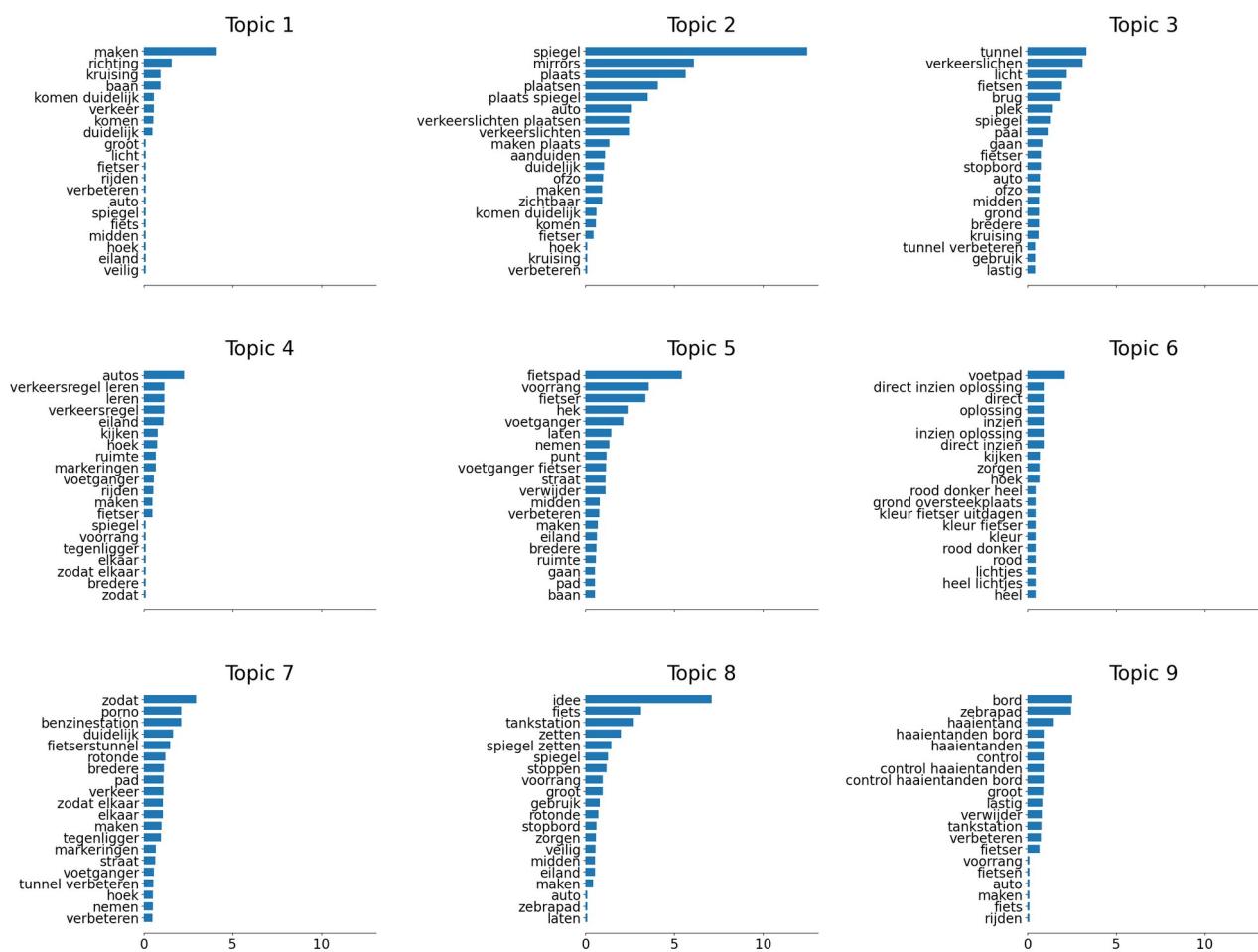
**Figuur 24: Aantal antwoorden per topics van Vraag 5**

### 3.2.2.1 Evaluatie van de topics



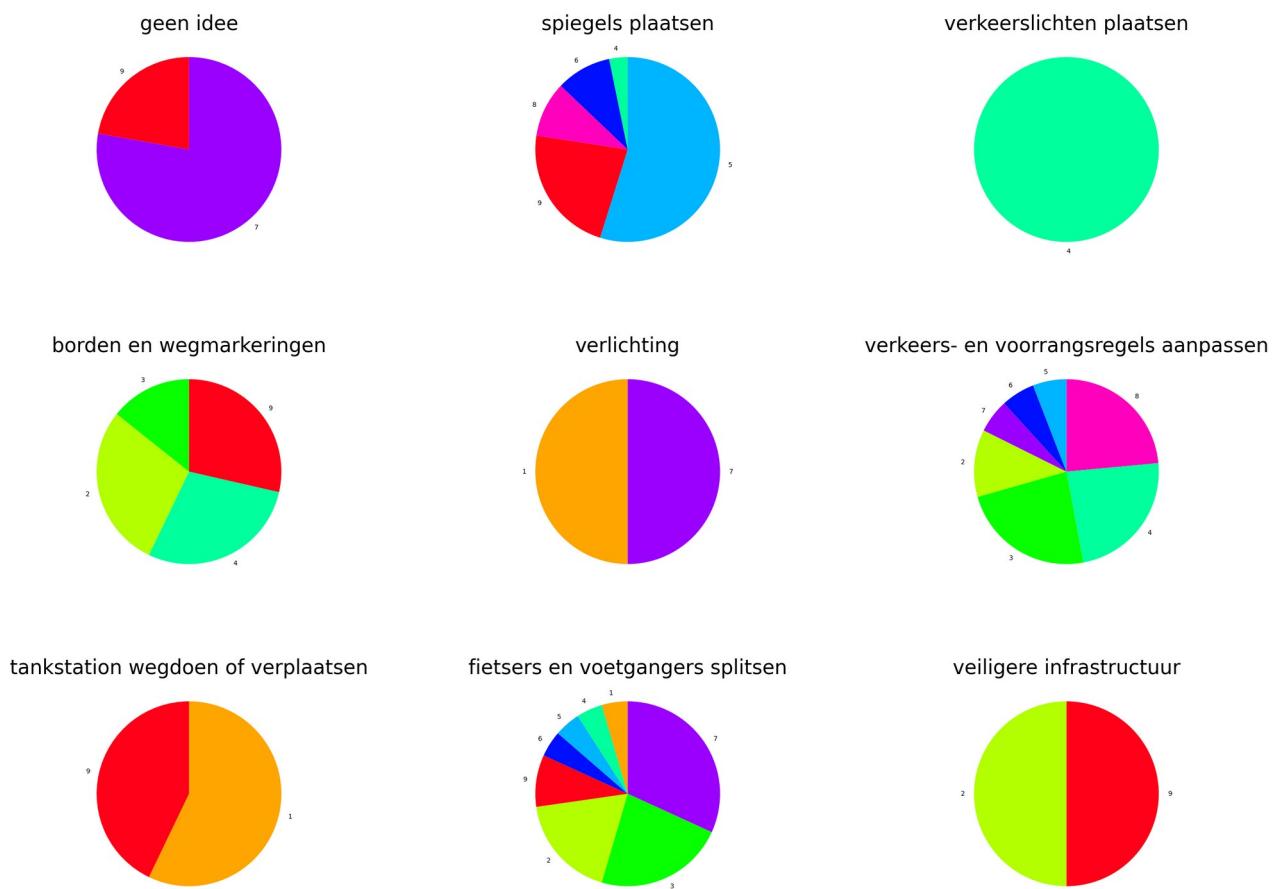
**Figuur 25: MDS plot van de topics van Vraag 5**

Het MDS-plot van de topics van Vraag 5 (Figuur 25) heeft nog steeds een grote stress factor, maar de topic clusters zijn duidelijker onderscheiden dan bij Vraag 4. “Spiegels plaatsen”, “verkeerslichten plaatsen”, “verlichting”, en “geen idee” zijn duidelijk geclusterd. “Tankstation”, “fietsers en voetgangers splitsen”, en “verkeers- en voorrangsregels aanpassen” zijn door elkaar geclusterd. De topics van Vraag 5 zijn dus al veel duidelijker dan de topics van Vraag 4.

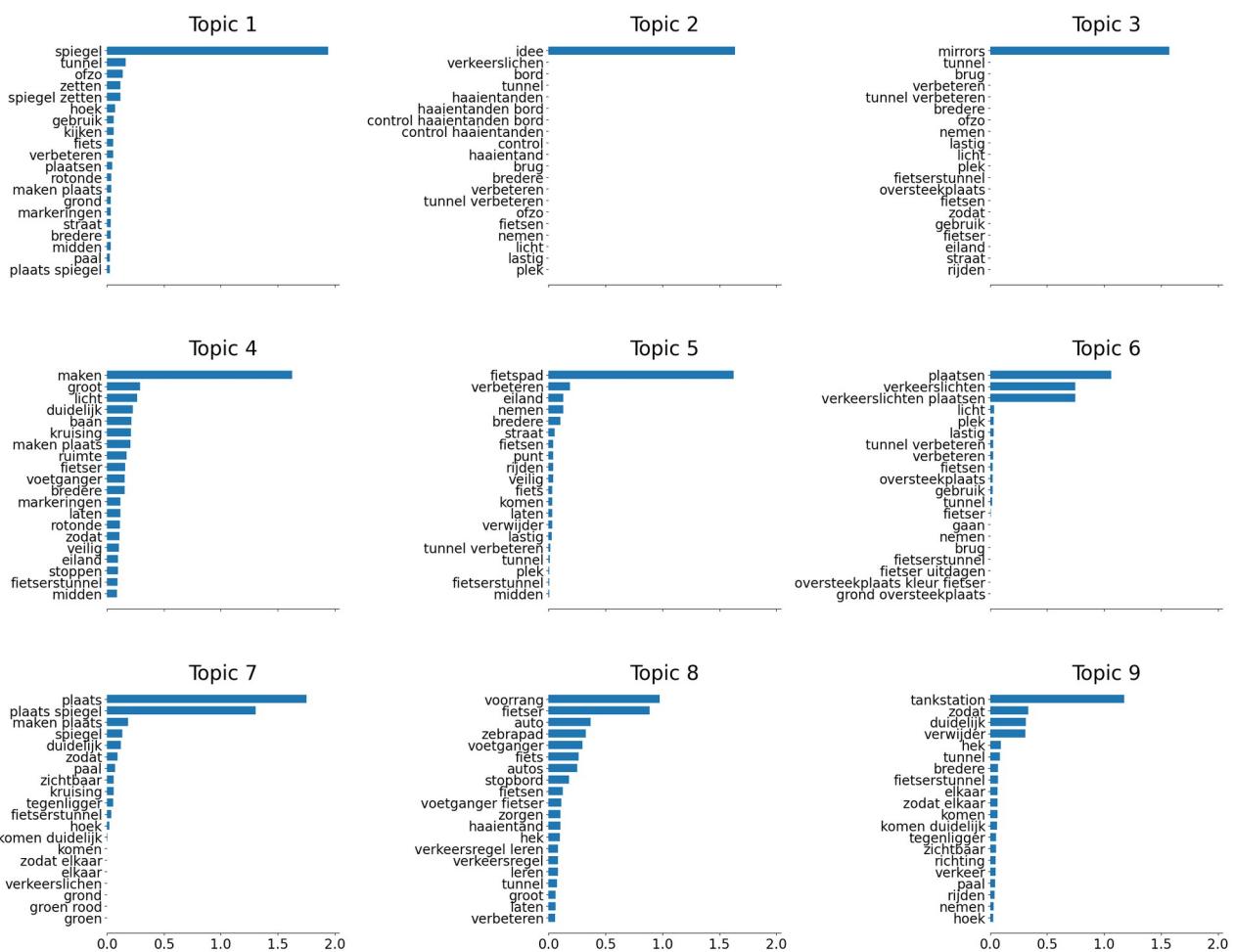


**Figuur 26: LDA topics voor Vraag 5**

Figuur 26 toont de LDA-topics voor Vraag 5. Hier zijn we uitgegaan van negen topics, aangezien we negen topics hebben geïdentificeerd met de hiërarchische clustering. Het is duidelijk dat topic 2 over het plaatsen van spiegels gaat, maar de andere topics zijn niet zo goed te identificeren. Figuur 27 toont de LDA-topics in vergelijking met de HC-topics voor Vraag 5. De HC-topics zijn niet zo sterk verspreid over de LDA-topics als bij Vraag 4. Vaak is een HC-topic verdeeld over 2 LDA-topics. Als we naar de kleuren overeenkomsten kijken kan je misschien zeggen dat “spiegels plaatsen” een verband heeft met “verkeerslichten plaatsen”.

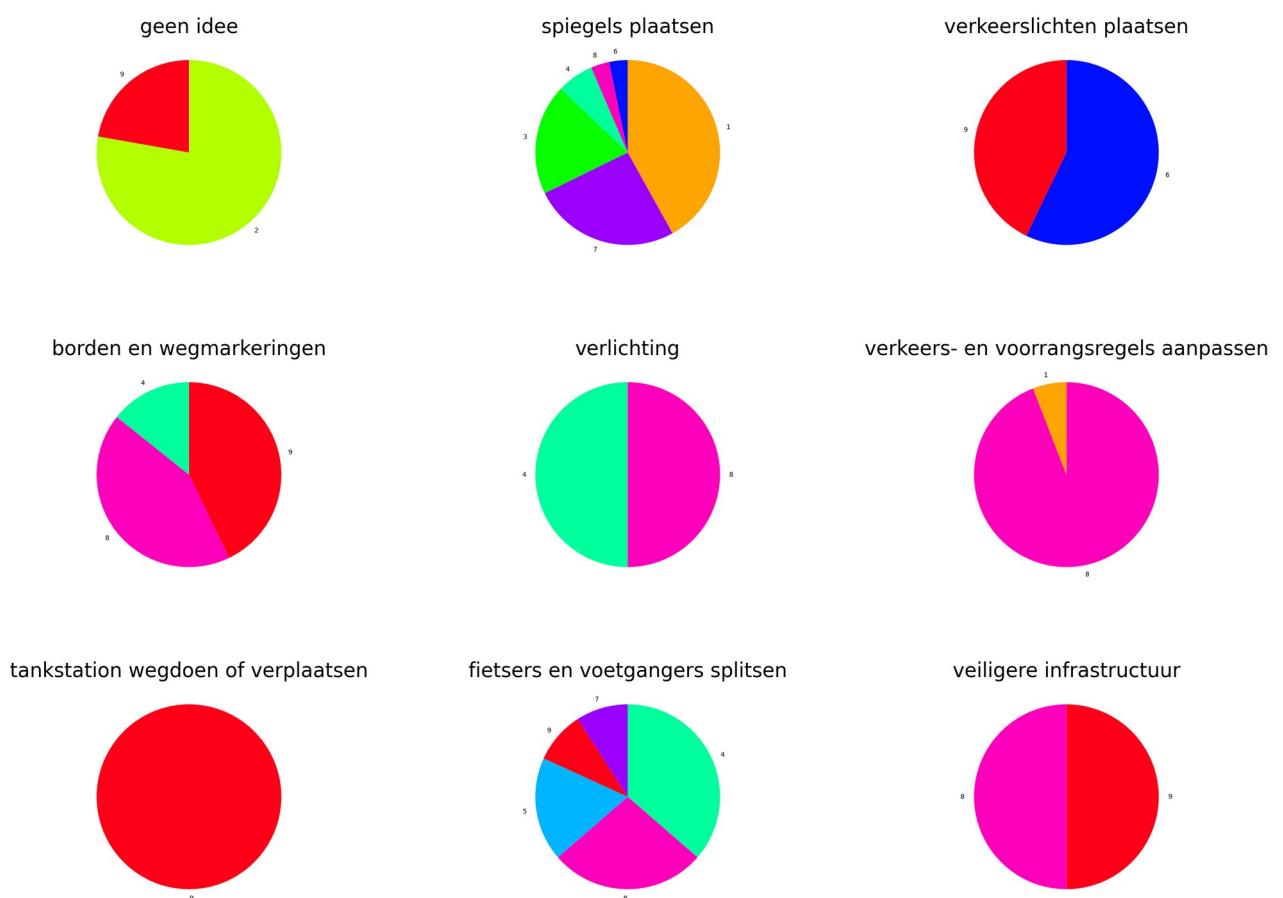


**Figuur 27: Vergelijking LDA topics per HC topic voor Vraag 5**



**Figuur 28: NMF topics voor Vraag 5**

Figuur 28 toont de NMF-topics voor Vraag 5. Deze zijn al wat duidelijker dan de LDA-topics. Topic 1 en 3 gaan bijvoorbeeld over de spiegels. Topic 2 over het feit dat de respondent geen idee had. Topic 6 gaat over het plaatsen van verkeerslichten en Topic 9 over het tankstation. De vergelijking met de HC-topics in Figuur 29 toont een sterkere homogeniteit. Het HC-topic over het tankstation bestaat volledig uit topic 9 (rood) en topic 8 (roze) toont de verbanden tussen “borden- en wegmarkeringen”, “verlichting”, “verkeers- en voorrangsregels”, en “veiligere infrastructuur”. Deze gaan allemaal over hoe een toevoeging van een hulpmiddel - licht, signalisatie - de veiligheid kan verbeteren.



**Figuur 29: Vergelijking NMF topics met HC topic voor Vraag 5**

### 3.3 Voorbeelden uit de LIManalysis tool

In dit deel geven we een paar voorbeelden over hoe de LIManalysis tool kan gebruikt worden om de data te verkennen, interpreteren, en zo inzichten te verzamelen. Met deze demo's proberen we de sterktes van de LIM-methode aan te tonen op een praktische basis.

#### 3.3.1 Verkeersveiligheid voor zwakke weggebruikers

De zwakke weggebruikers - fiets, te voet, met de skate of step - is de grootste groep onder de respondenten. De gevaarlijkste plaatsen voor zwakke weggebruikers op de Sterre zijn volgens Figuur 30 de fietserstunnel en de zijgang van de campus De Sterre waar de hekjes staan. Als onderzoeker kan

je dan geïnteresseerd zijn naar wat deze hotspots gevaarlijk maakt. Door te klikken op de fietserstunnel, kan je de topics zien die daar op toepassing zijn. Dit zijn zicht, gevaarlijke plaats, gevaarlijke weggebruikers, veel weggebruikers, snelheid, voorrangsregels, gebrek infrastructuur, en steile helling. Dit kan je interpreteren dat de weggebruikers door de steile helling daar snelheid opbouwen, maar niet zien wat er achter de bocht is in de tunnel. Dit kan leiden tot gevaarlijke situaties. De rechter wordcloud in Figuur 31 toont wat er kan gedaan worden om de verkeersveiligheid te verbeteren daar, namelijk het plaatsen van spiegels. Dit klinkt een goed idee. Spiegels zullen er namelijk voor zorgen dat mensen kunnen zien wat ervan achter de bocht komt, en kunnen zo dan beter het risico inschatten en hun snelheid proberen te minderen. Voor de tweede hotspot gaat het vooral over de aanwezigheid van hekjes die in de weg staan, en over hoe het onduidelijk is over wat de voorrangsregels zijn bij het oversteken van de baan. Er zijn veel te veel verschillende redenen daar waardoor het moeilijk is om één duidelijk reden te vinden waarom het gevaarlijk is. Het is daar dus gevaarlijk door een mengeling van verschillende redenen en door de onnauwkeurigheid bij het aanduiden van de gevaarlijke gebieden, zijn de verschillende redenen moeilijk ruimtelijk te splitsen.



**Figuur 30: Heatmap van gevarenzones voor de zwakke weggebruikers**

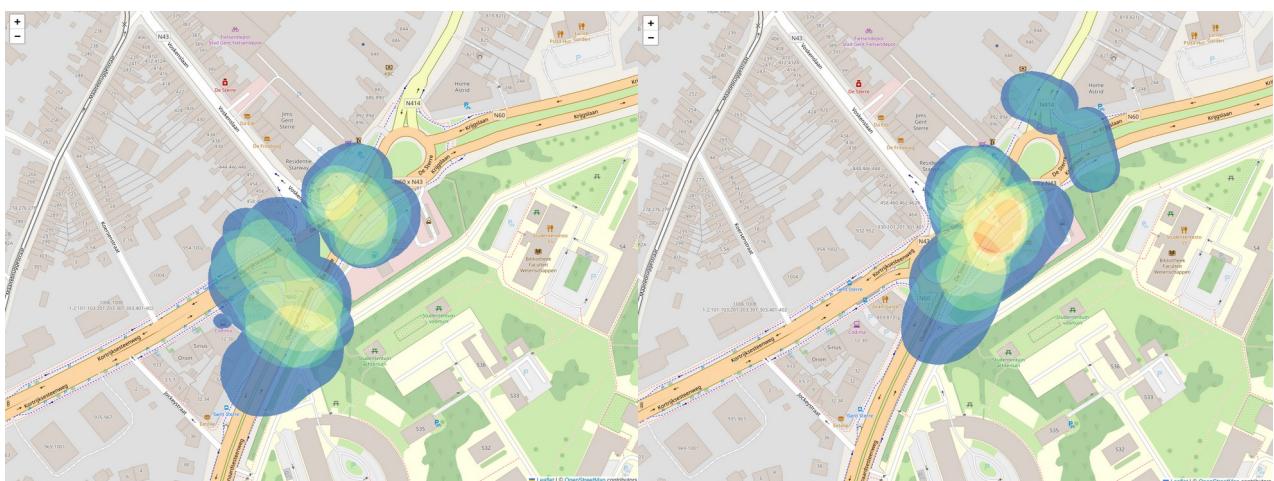


**Figuur 31: Wordclouds zwakke weggebruikers in de fietserstunnel.**  
**Links: beschrijving. Rechts: verbetering**

### 3.3.2 Veiligheid onderweg naar de S8

Een onderzoeker kan bijvoorbeeld juist geïnteresseerd zijn over de het gedrag en verkeersveiligheid van het verkeer dat naar de S8 gaat. Deze kan dan filteren op OD-zones met de OD-zone van de S8 als bestemming. Door dan te gaan over de andere zones als oorsprong, kan deze onderzoeker uitzoeken welke route er als gevaarlijk wordt aanschouwd. Er zijn twee oorsprongen die vaak werden beoordeeld met als bestemming de S8. Dit zijn de Voskenslaan en de Kortrijksesteenweg richting het station. Figuur 32 toont de gevarenzones voor deze met als origin de Voskenslaan (links), of de Kortrijksesteenweg (rechts). Beide hebben veel de fietserstunnel aangeduid, en ook twee oversteekplaatsen. Bij de Voskenslaan is dit aan de westkant van het rondpunt, en bij de Kortrijksesteenweg is dit aan de oostkant van het rondpunt waar de oorsprong dichtbij ligt. Voor de fietserstunnel gaat het voor beide oorsprongen over de gevaarlijke weggebruikers en het zicht achter de hoek. Bij de oversteekpunten van de Voskenslaan gaat het vooral over het

vele autoverkeer en dat ze de voorrangsregels niet goed kennen en bij de oversteekpunten van de Kortrijksesteenweg gaat het eerder over de fietsers in tegenrichting. Bij de Voskenslaan stelt men dan ook eerder voor om de oversteek te verbeteren of de voorrangsregel duidelijker maken, terwijl bij de Kortrijksesteenweg niets van verbeteringen worden voorgesteld naast spiegels plaatsen. Zo kan de onderzoeker besluiten dat vooral de fietserstunnel veiliger moet gemaakt worden door het plaatsen van spiegels, of dat er een veiligere oversteek moet komen aan de westkant van het rondpunt. De oostkant is al veilig genoeg.

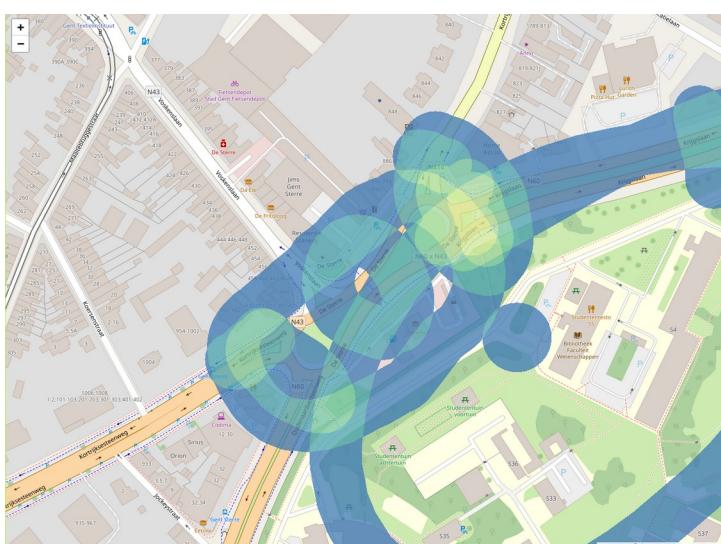


**Figuur 32: Gevarenzones voor routes met als bestemming de S8 en als oorsprong de Voskenslaan (links) of de Kortrijksesteenweg (rechts)**

### 3.3.3 Houd uw snelheid

Als onderzoeker kan je geïnteresseerd zijn naar de plaatsen waar mensen te snel rijden. Eén van de topics die we uit de antwoorden op de open vragen hebben kunnen halen is snelheid. De LIManalysis tool maakt het mogelijk om te filteren op topic en dus enkel de gevarenzones te tonen die als topic snelheid hebben. Figuur 33 toont het resultaat van deze filtering. Eén zone valt direct op, namelijk de oversteekplaats van de Krijgsstraat aan de oostkant van het rondpunt. Bij het verder inspecteren van deze plaats leren we in de

Wordcloud op Figuur 34 dat dit gaat over het abrupt remmen en omdat er files zijn. Als een onderzoeker deze plaats heeft geïdentificeerd als een plaats waar auto's altijd te snel rijden, kan deze op zoek gaan naar manieren om deze snelheid te verminderen. Eén manier dat is voorgesteld door de participanten is een stopbord, maar dit kan enkel bij het oprijden van het rondpunt, niet bij het afrijden waarbij het sneller gaat, en een stopbord is niet naar gewoonte bij een rondpunt. Een andere manier dat wij voorstellen is een snelheidsdrempel bij het zebrapad.



**Figuur 33: Heatmap van waar mensen praten over de snelheid van de verkeer**

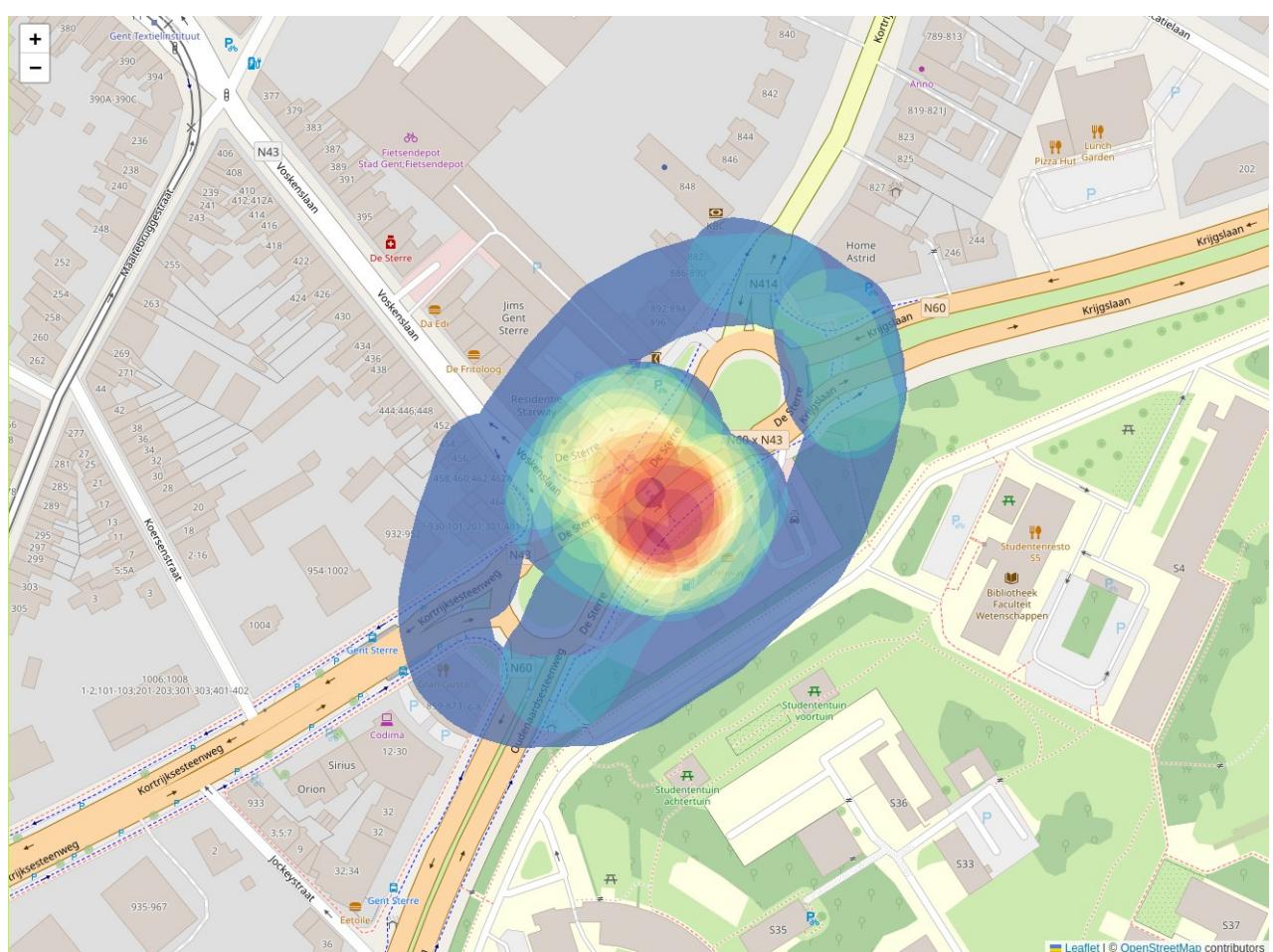


**Figuur 34: Wordcloud van de snelheid aan de Krijgslaan**

### 3.3.4 Waar zijn er spiegels nodig

We hebben al geïdentificeerd dat spiegels een zeer vaak voorgestelde verbetering is voor bij het rondpunt. De stad kan daarbij dan bijvoorbeeld vragen waar het interessantste is om spiegels te plaatsen. Door te filteren op de verbetering “spiegels”, krijg je Figuur 35 dat toont welke zones er als gevaarlijk werden aangeduid waarbij de participant ook spiegels voorstelde als verbetering. Het is duidelijk dat hierbij vooral de fietserstunnel wordt

aangeduid. Daar wordt ook vaak gezegd dat er geen zicht is achter de hoek. Opvallend is wel dat het vooral de zuidkant van de fietserstunnel is, aan de kant van het tankstation, dat wordt aangeduid. Aan de noordkant, bij de supermarkt Delhaize, is de zichtbaarheid dus nog goed genoeg, en is het vooral bij de zuidkant dat mensen te veel moeten opletten. Door een spiegel te plaatsen aan de zuidkant van de fietstunnel, kan Stad Gent de verkeerssituatie op de Sterre een heel stuk veiliger maken.

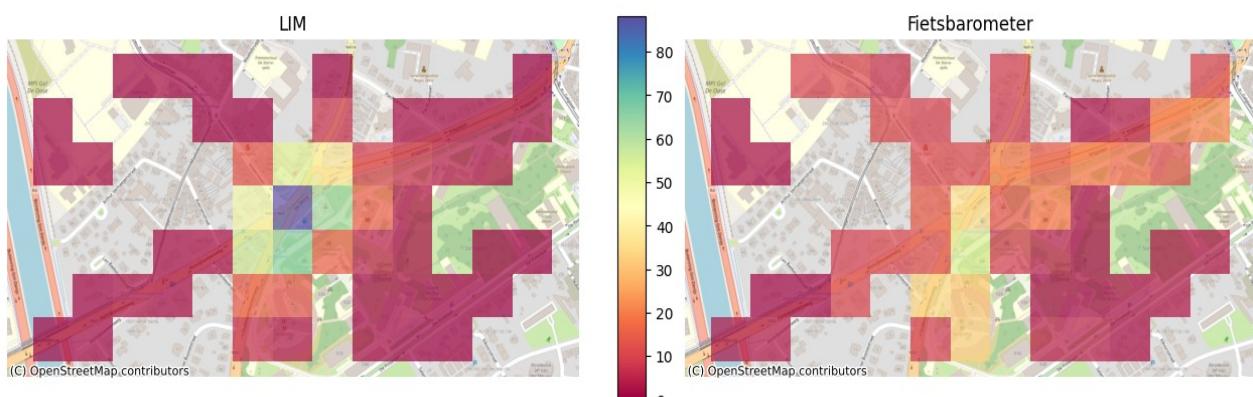


**Figuur 35: Heatmap waar er vaak om spiegels worden gevraagd**

## 3.4 Vergelijking met de Fietsbarometer

### 3.4.1 Quadrant Analysis met een $\chi^2$ Goodness-Of-Fit test

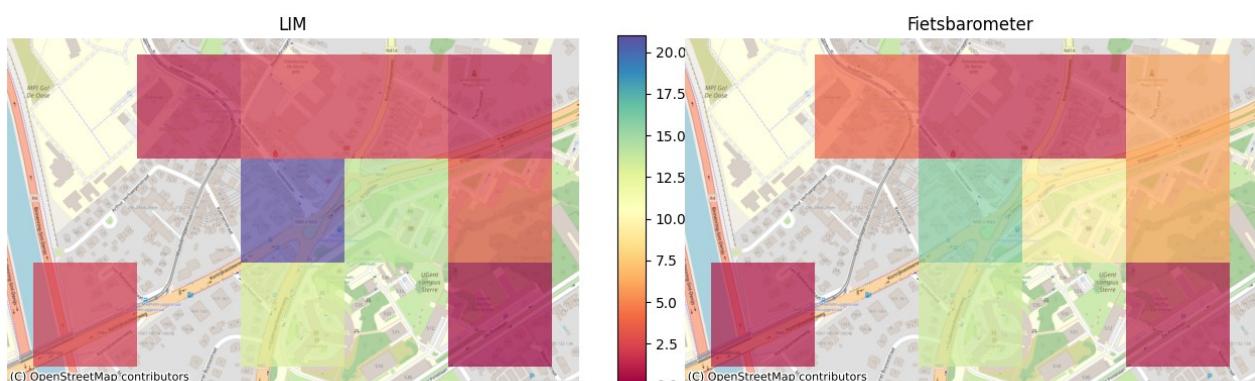
Figuur 36 toont de resultaten van de QA voor de volledige dataset voor de Sterre. De kleurenschaal toont het aantal herschaalde observaties er per kwadrant zijn. Beide methoden tonen ongeveer dezelfde vorm, maar de LIM methode is meer geconcentreerd op het midden van het rondpunt, terwijl de Fietsbarometer meer verspreid is. Bij de Fietsbarometer werd er ook gevraagd om wegsegmenten te beoordelen, die redelijk langwerpig kunnen zijn en zich verspreiden over meerdere quadranten. De  $\chi^2$  Goodness-Of-Fit test heeft een waarde van 915,2 met 51 degrees of freedom, wat zorgt voor een p-waarde van 0,0. We verwerpen dus de nulhypothese dat beide datasets dezelfde ruimtelijke distributie hebben.



**Figuur 36: QA vergelijking tussen LIM en Fietsbarometer**

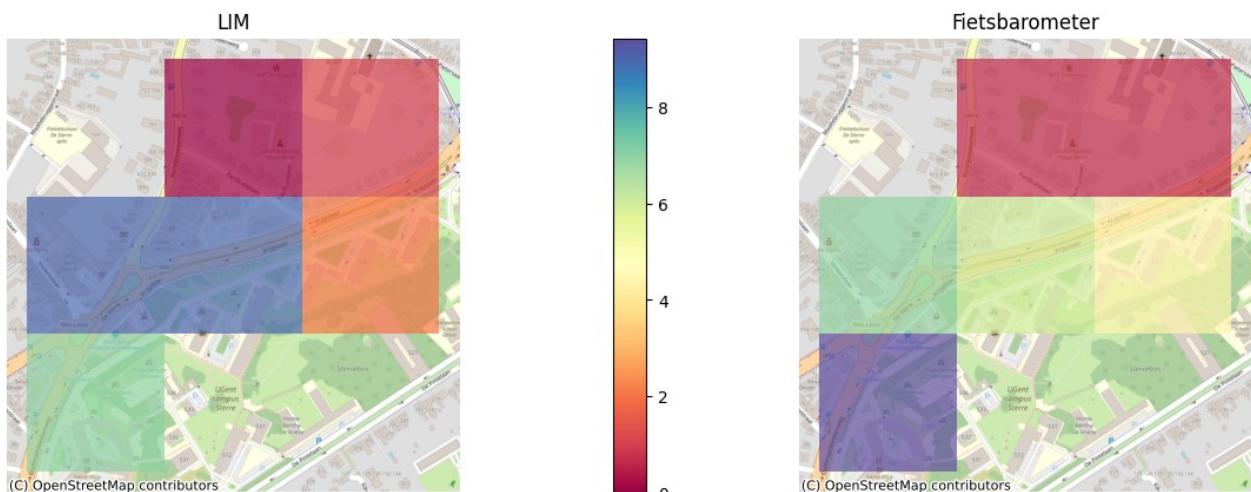
De topics drukte en snelheid die uit de NLP hiërarchische clustering kwamen in hoofdstuk 3.2.1, werden ook onderzocht in de Fietsbarometer. Dus lijkt het interessant om ook een QA te doen voor juist de observaties waar deze topics belangrijk zijn. Figuur 37 toont de QA voor het topic drukte. Omdat we enkel de observaties bekijken waar dit topic belangrijk is, zijn er minder observaties en zal Griffith et al.'s (1990) formule voor grotere cellen zorgen. De

Fietsbarometer is nog steeds meer verspreid dan de LIM-methode, maar ze lijken al wat meer op elkaar. De  $\chi^2$  Goodness-Of-Fit test toont ook meer gelijkenis met een  $\chi^2$  waarde van 14,53 met 9 degrees of freedom, wat zorgt voor een p-waarde van iets meer dan 0,1. We aanvaarden dus onze nulhypothese dat de twee datasets ruimtelijk gelijk verspreid zijn op vlak van drukte.



**Figuur 37: QA vergelijking voor drukte**

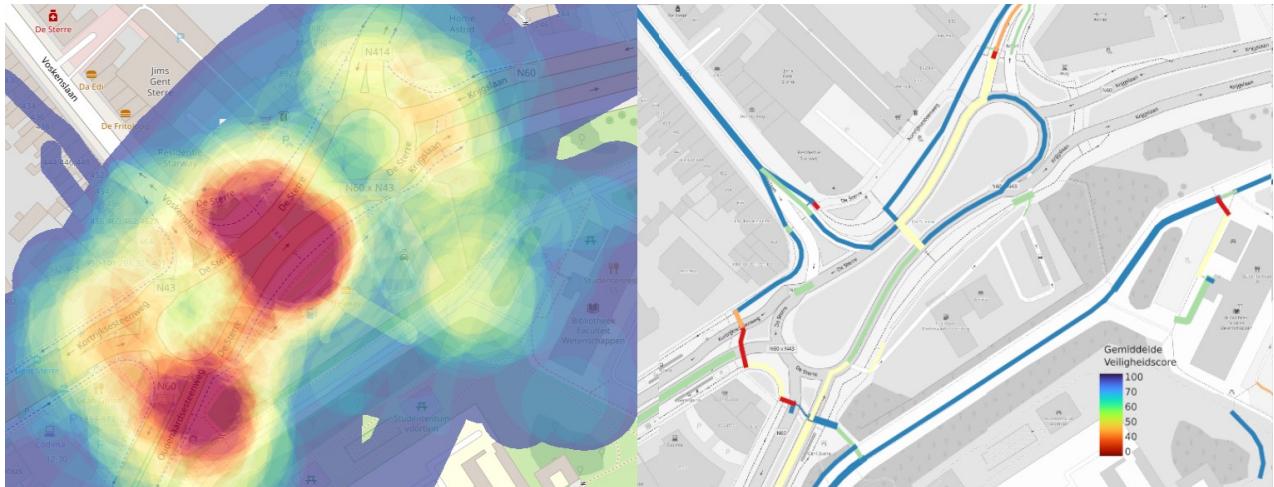
Figuur 38 toont de QA van alle observaties waarbij het topic snelheid belangrijk zijn. Er zijn nog maar zeer weinig grote cellen waardoor er details verloren kunnen zijn. Visueel lijken ze sterk op elkaar. De  $\chi^2$  Goodness-Of-Fit test, heeft een waarde van 4,9428 met 5 degrees of freedom wat zorgt voor een p-waarde van 0,42. We aanvaarden dus onze nulhypothese dat de twee datasets een gelijke ruimtelijke verdeling hebben van plaatsen waar het topic “snelheid” belangrijk is.



**Figuur 38: QA vergelijking voor snelheid**

### 3.4.2 Visuele vergelijking tussen LIM en de Fietsbarometer

Figuur 39 toont hetzelfde gebied met de data van de LIM-methode (links) en de data van de Fietsbarometer (rechts). De kaart van de LIM-methode toont een heatmap van hoe vaak een bepaald gebied is aangeduid zoals gezien in 3.3. De kaart van de Fietsbarometer toont de gemiddelde veiligheidsscore van een wegsegment als kleur, en het aantal beoordelingen van een wegsegment is logaritmisch evenredig met de dikte van de lijn. Het is dus belangrijk om te begrijpen dat beide figuren niet hetzelfde voorstellen. Bij de Fietsbarometer is er een score dat zegt hoe gevaarlijk mensen het daar vinden op een schaal van zeer gevaarlijk (0) tot zeer veilig (100). Terwijl er bij de LIM-methode nu enkel werd gevraagd naar waar men het gevaarlijk vindt, en enkel kan aantonen hoe vaak mensen een gebied hebben aangeduid als gevaarlijk, zonder dat er een impliciete manier is om te zeggen dat een gebied veilig is. De resultaten van de Fietsbarometer zijn dus al completer in de zin dat het ook rekening houdt met gebieden dat veilig bevonden worden.



**Figuur 39: Vergelijking LIM (links) en Fietsbarometer (rechts)**

Bij de Fietsbarometer is de respondent beperkt tot het volledige wegsegment in het geheel, of meerde wegsegmenten tezamen te beoordelen. Het is er niet mogelijk om juist één bepaald deel van een wegsegment aan te duiden. Ook is de respondent beperkt tot de wegsegmenten die door de onderzoeker gegeven zijn te beoordelen. Ze hebben dus niet de mogelijkheid om een arbitrair gebied op het publiek domein aan te duiden en te beoordelen. De LIM-methode laat toe om eender welk gebied aan te duiden en te beoordelen, en is hierbij niet beperkt tot enkel maar straten en volledige wegsegmenten. De LIM-methode is dus beter in het vangen van details.

Ramírez Aranda et al. (2021) ondervonden al dat een marker accurater er is dan een lijn, of punt of polygoon. Ook al is de lijn van de Fietsbarometer niet ingetekend door de respondent, toch zien wij dat vaak de grote baan is aangeduid voor fietsers in plaats van het fietspad. Bij de LIM-methode zien we dat vooral de fietstunnel is aangeduid als gevaarlijk, als ook alle oversteekplaatsen, de zijingang van campus de Sterre, en de ingang van het benzinstation. Bij de Fietsbarometer is het vooral een paar oversteekplaatsen die rood kleuren en de fietserstunnel die als licht gevaarlijk wordt aanschouwd. De LIM-methode is ruimtelijk accurater dan de Fietsbarometer, maar is minder precies door de grote marker.

Eigenschappelijk is de Fietsbarometer veel uitgebreider en beter gedefinieerd. De LIM-methode is attribuutaal minder nauwkeurig doordat veel diverse antwoorden samen worden geclusterd om er gedefinieerde onderwerpen uit te krijgen.

Storme et al. (2022) vonden dat ze al veel inzichten kunnen trekken uit de Fietsbarometer. Bijvoorbeeld dat een apart fietspad veiliger wordt aanzien dan een fietspad naast een straat waar er auto's rijden. Ook konden ze zien welke intersecties als veilig of als gevaarlijk werden aanzien. Met de LIM-methode kan je makkelijk zien waar de meeste mensen het gevaarlijk vinden, en ook waarom ze het daar gevaarlijk vinden. Ook is het verkennen via de analyse tool makkelijk in gebruik om de resultaten te interpreteren en er conclusies uit te trekken. Eén van de problemen dat Storme et al aanhielden was dat er een hoge lokale participatiegraad nodig is om er significante inzichten uit te halen. Met de LIM-methode was het doel om een hoge participatiegraad te behalen door de vragenlijst kort en simpel te houden, maar zijn we erin geslaagd om met maar 174 participanten te halen. Toch konden we met dit aantal participanten toch enorm veel inzichten achterhalen over een klein gebied. Zowel de Fietsbarometer als de LIM-methode zorgen dus voor veel en goede inzichten, maar ze geven beiden verschillende inzichten en kunnen elkaar dus verrijken. De LIM-methode is beter om exploratief onderzoek te doen en om een gevoel te krijgen hoe mensen algemeen over een thema denken en wat hun erin bezighoudt. Voor een gerichtere studie te doen over een bepaald fenomeen is een uitgebreider onderzoek beter.

## 4 DISCUSSIE

In deze studie hebben we ondervonden dat de LIM-methode een nuttige manier is om geografische data te verzamelen bij de gewone mens, zonder ze te vragen om al te veel moeite te doen. De verkregen data zit vol met

inzichten en lokale kennis, en het koste niet veel moeite om deze kennis te extraheren en samen te vatten zodat het makkelijk allemaal te analyseren en interpreteren was. De LIM-methode heeft haar voordelen en nadelen en heeft haar eigen sterkes en zwaktes in vergelijking met andere projecten om geografische data te verzamelen bij de mensen zoals de Fietsbarometer (Storme et al., 2022). Toch zijn er een aantal zaken die we post hoc pas ondervonden dat beter kan.

#### **4.1 Dataverzameling**

Ook al was er al genoeg data om interessante conclusies uit te trekken, kon de dataverzameling veel beter verlopen zodat er meer antwoorden komen van een divers publiek. Momenteel kwamen de meeste antwoorden van de opstelling met het groot touchscreen aan de inkom van de S8. Het is echter niet altijd mogelijk om zo'n opstelling op te zetten. Ook kwam daardoor de meeste antwoorden van mensen die moeten zijn in de S8, namelijk studenten en onderzoekers die werken aan de Universiteit Gent. De hoop van de LIM-methode is om met zo met mogelijk zo veel mogelijk te behalen, dus ook een groot en gevarieerd doelpubliek. Daarom zou er in de toekomst betere reclame moeten gemaakt worden zodat alle mensen die bijvoorbeeld langs het rondpunt passeren de kans krijgen om de vragenlijst in te vullen, en niet enkel studenten.

Het ophangen van flyers en posters rond het rondpunt was geen effectieve strategie. Mensen kijken niet meer zo veel naar uitgehangen reclame, en de meeste zelf uitgehangen reclame blijven niet lang hangen. Wij wouden te sterk inzetten op posters met een QR-code zodat mensen dit dan lokaal zouden scannen en de vragenlijst aan het studiegebied zelf invullen. Dit zorgde voor veel teleur gesteldheid toen dat er amper respons kwam op de posters. Jankowski et al. (2016) onderzochten de voorkeuren van mensen over het invullen van geo-questionnaires waarbij mensen werden uitgenodigd om

hun mening te brengen over landgebruik in hun regio via een online Web gebaseerde tool. Zo vonden dat de meeste respondenten op hun tool zijn gekomen via advertenties in lokale media, gevolgd door advertenties in sociale media zoals Facebook, en ook veel via post invitatie en mondelinge reclame door vrienden. Maar vijf van hun duizend respondenten kwamen via posters.

Voor toekomstig onderzoek is het dus zeer interessant om op andere manieren reclame te maken. Een eerste andere mogelijk manier om reclame te maken is om massaal flyers te steken in brievenbussen van buurtbewoners. Bij vele mensen zal dit dan nog direct in de vuilnisbak belanden, maar als we bij genoeg buurtbewoners reclame maken voor de korte vragenlijst, zullen er toch nog een heel deel de vragenlijst invullen. Hiervoor moeten er dan wel veel flyers afgeprint worden, wat veel papier en geld kost, en veel werk is om al deze flyers in brievenbussen te steken.

Een tweede manier is om online te adverteren op Facebook of andere sociale of lokale media. Dit komt overeen met een deel van het onderzoeker overlaten aan derde partijen. Dit is dan minder werk voor de onderzoeker en je kan ook meer mensen bereiken, maar er hangt er een hogere kost aan vast terwijl de LIM-methode erom gaat om het met minder te doen. Online adverteren op sociale media is echter niet zo duur, en met gerichte advertenties kan je makkelijk je doelpubliek bereiken. Ook kan je in lokale Facebookgroepen of andere internetfora gratis reclame maken voor je onderzoek. Het nadeel is dan wel dat mensen niet meer ter plaatse zijn om het studiegebied te beoordelen.

Een derde manier om mensen aan te zetten om de vragenlijst in te vullen, is om lokaal mensen aan te spreken of ze vijf minuten van hun tijd willen doneren om het in te vullen. Dit heeft al het voordeel dat mensen ter plaatse zijn en zo beter gekend zijn met de lokale situatie waardoor ze beter kunnen

oordelen. Ook zijn mensen eerder geneigd om mee te doen aan een onderzoek of een vragenlijst in te vullen als dit hun persoonlijk wordt gevraagd (Roghanizad & Bohns, 2017). Een nadeel is dat de onderzoeker veel mensen moet aanspreken zodat ze de vragenlijst invullen, wat dan weer tegengaat met het *less*-aspect van *Less Is More*.

Bij het opstellen van open vragen is het belangrijk dat de vraag duidelijk is en dat de respondent zonder hulp begrijpt wat de onderzoeker bedoeld met die vraag (Züll, 2016). Als de vraag verkeerd begrepen is door de respondent, zal deze een antwoord geven op een andere vraag dan de onderzoeker had gesteld waardoor het moeilijker wordt voor de onderzoeker om de antwoorden te analyseren. Bij de vijfde vraag, “Hoe zou u dit verbeteren?”, hadden wij een paar respondenten die dachten dat dit ging over de vragenlijst zelf. Deze hadden dan ook feedback gegeven over de vragenlijst in plaats van suggesties te geven over het verbeteren van de verkeerssituatie. Alhoewel het altijd interessant is om feedback te krijgen, was er hier al een veld voor voorzien op het einde van de vragenlijst.

Bij het tekenen op de kaart heeft de respondent ook alle vrijheid wat die tekent en is hierbij niet gelimiteerd tot een bepaald stratenennetwerk of voor gedefinieerde regio's. Met deze analogie kan je dus zeggen dat het tekenen op de kaart ook een soort open vraag is, waardoor het aantal echt gesloten vragen in de vragenlijst wordt gereduceerd naar één. De LIM-methode probeert dus haar resultaat te halen bijna volledig met behulp van open vragen. We hadden wel net aangehaald dat het belangrijk is om de vraag zeer goed te formuleren bij het opstellen van open vragen en dit is niet anders bij het tekenen op de kaart.

Bij vraag 3, “duid één plaats aan waar op dit kruispunt u het gevaarlijk vindt met uw gekozen vervoersmiddel”, hebben we dit proberen zo goed mogelijk te formuleren zodat de participanten: maar één plaats aanduiden, en dit

aanduiden door een regio volledig in te kleuren. Sommige participanten hadden meerde plaatsen aangeduid. Er stond duidelijk in de vraag dat ze maar één plaats mogen aanduiden, maar anders werd dit niet geforceerd dat ze maar daadwerkelijk maar één plaats aanduiden. Het werd zelfs toegelaten doordat het tekenen op de kaart werkt met aan polylijn in plaats van een enkele lijn. De polylijn werd gekozen zodat de participanten het makkelijker hadden om meer complexere regio's aan te duiden, maar er is nergens een check toegevoegd dat de getekende polylijnen één gebied vormen. Dit heeft een effect op de resultaten omdat de reden waarom het gevaarlijk is wordt gelinkt aan de aangeduide zones. De participant kan maar één keer beschrijven waarom het gevaarlijk is, maar beoordeeld daarmee soms ongewenst alle zones die ze hebben aangeduid.

Ook de tweede verwachting bij het aanduiden van de zone werd soms niet gevuld. Participanten duidden soms een gebied aan door er een cirkel rond te trekken in plaats van het gebied in te kleuren. Dit staat echter niet duidelijk geformuleerd in de vraagstelling. "Duid aan" is een dubbelzinnige opdracht die op verschillende manier kan geïnterpreteerd worden. De onderzoekers verwachten dat de respondent dan een gebied inkleurt, maar de respondent kan hierbij ook een gebied omcirkelen, onderlijnen, arceren, etc. ondervonden ditzelfde probleem bij hun eerste iteratie van *My Green Place* en probeerden dit te vermijden door aan de respondent te vragen om het gebied "te schilderen", vooraf een instructievideo te tonen hoe ze een gebied moeten schilderen, en tegen het einde van het aanduiden nog eens een bericht tonen om enkel lege plaatsen te laten als ze dit daadwerkelijk zo bedoelden. Wij stellen ook voor om ook te controleren of het gebied dat de respondent aanduidt geen gaten of ringen bevat, tenzij het mogelijk moet zijn dat er gaten midden in een aangeduid gebied zit.

Bij vraag twee, "teken de route die u volgt als u zich over dit kruispunt verplaatst met uw gekozen vervoersmiddel", waren de verwachtingen dat de

respondenten een aaneengesloten lijn tekent in de richting dat ze deze route ook afleggen. Opnieuw was het wel mogelijk voor de respondent om meerdere lijnen te tekenen met het idee om het makkelijker te maken om complexere routes te tekenen, maar er werd niet gecontroleerd of deze lijnen elkaar opvolgden. Het enige wat werd gedaan was de verschillende getekende lijnen verbinden met elkaar bij het verwerken van de data. Respondenten hebben dan ook wel routes getekend bestaande uit meerdere lijnen die niet elkaar opvolgen. Ook waren er inzendingen met rare uitschieters omdat er bij het intekenen iets raars is gebeurd, bijvoorbeeld een extra onverwachte vinger op het scherm. Er was ook een inzending waarbij de respondent pijlen had getekend op de route om de richting aan te duiden. Dit zou niet zozeer kunnen vermeden worden door een betere vraagstelling, maar misschien wel door een betere representatie van de route als de respondent deze aan het inteken is. We zouden kunnen de verschillende direct verbinden zodat de respondent ziet dat het een aaneengesloten route is. Ook zouden we kunnen controleren op plotse uitschieters en deze dan blokkeren of een nieuwe route starten als de uitschieter doorgaat. Een indicatie van richting, met bijvoorbeeld pijlen in de lijn, kan de respondent een hint geven dat de route dat ze intekenen een richting heeft.

De respondenten hadden ook de mogelijkheid om feedback te geven op de vragenlijst. De meest voorkomende feedback was dat het groot touchscreen moeilijk in gebruik was, en dat het moeilijk was om het digitaal toetsenbord te gebruiken op het groot touchscreen. Het groot touchscreen was dan ook pas een later idee, en de vragenlijst was niet ontworpen om te gebruiken op zo'n groot scherm. Alhoewel het dus de aandacht trekt van de mensen en zo zorgt voor veel respondenten, is het groot touchscreen moeilijker te gebruiken en zorgt het dus voor een lagere kwaliteit van ingegeven data. Een ander puntje feedback was dat de UI niet zo mooi is, en dat toont dus aan dat het belangrijk is een goeie UI te hebben zodat mensen niet worden afgeschrikt van het invullen van de vragenlijst.

## 4.2 Data-Analyse

Bij Natural Language processing zit de kracht van verschillende technieken soms al bij een goede pre-processing. Bij het hiërarchisch clusteren bijvoorbeeld, waren er vaak clusters die bestonden uit twee of drie keer het exactzelfde antwoord. Door een technische fout is het dus soms gebeurd dat een inzending meerdere keren is doorgestuurd. Of het kan ook zijn dat dat iemand meerdere inzendingen heeft gedaan waarbij die verschillende zones aanduidt maar telkens hetzelfde antwoord meegeeft op de open vragen. Bij het eerste geval moeten die dubbele antwoorden dus weggehaald worden. Bij het tweede geval kan dit voor zeer sterk gedefinieerde miniclusters zorgen, waardoor deze niet zo snel worden samengevoegd met andere antwoorden die over hetzelfde spreken. Duplicaten detecteren en weghalen houdt dus niet enkel in de antwoorden op de open vragen vergelijking, maar ook de ingetekende geometrieën te vergelijken. Dit is normaal zeer simpel als de geometrieën exact hetzelfde zijn. Als de geometrieën door een technisch probleem zeer lichtjes verschillen, bijvoorbeeld een punt dat 0,0001 eenheden is opgeschoven, is het moeilijker om te bepalen of twee geometrieën hetzelfde zijn en dus duplicaten zijn, maar normaalgezien als het duplicaten zijn door een dubbele inzending zijn de geometrieën exact hetzelfde.

Sommige antwoorden bestonden ook uit meerdere zinnen die verschillende redenen gaven. Dit zijn ofwel echte zinnen gescheiden door een punt, ofwel samengestelde zinnen gescheiden door een “en” of komma, ofwel zinnen gescheiden door een nieuwe regel. Het antwoord vak bestond ook uit meerdere regels, en mensen geven ook langere antwoorden als het antwoord vak groter is (Smyth et al., 2009). Sommige mensen hebben dus de meerdere regels die er waren om de tekstromloop te faciliteren gezien als een reden dat je meerdere antwoorden en meerdere redenen moet geven. Hoewel dit niet tegen het principe van de LIM-methode is – meer antwoorden – moet er hier toch rekening mee gehouden worden bij het verwerken van de antwoorden.

Als een antwoord met meerdere zin wordt geclusterd en geklassificeerd, zal deze maar één topic toegekend krijgen ook al gaat het over meerdere topics. De antwoorden zouden dus gesplitst moeten worden in hun zinnen. Dit laat dan wel de vraag of dat dan meerdere topics worden gehangen aan één inzendingen, of dat de inzending wordt geduplicateerd per topic waardoor je meerdere dezelfde geometrieën hebt. De eerste optie is complexer om te implementeren, maar de heatmap wordt er niet door beïnvloed, terwijl dit bij de tweede optie wel het geval is doordat je meerdere inzendingen creëert. Of misschien is dat wat gewenst is, aangezien er meerdere redenen worden gegeven waarom een regio gevvaarlijk is, kan het ook zijn dat deze zone dus meerdere keren moet aangeduid worden, één keer per reden.

Een negatie in een zin is meestal maar één woord, zoals “niet” of “nooit”, maar zorgt wel dat de zin over totaal het tegenovergestelde gaat. Bijvoorbeeld één van de inzendingen had als antwoord op vraag 4 “Ik vind het hier niet gevvaarlijk”. Dit antwoord wordt echter geclusterd bij het topic “gevaarlijke plaats”. Dit is een probleem aangezien het te makkelijk totaal verkeerd kan geïnterpreteerd worden door de clustering. Ergens moet dus rekening gehouden worden met negaties bij het analyseren van de open vragen. Morante & Blanco (2021) geven een overzicht over hoe men met negaties kan omgaan bij Natural Language procesing.

Een zeer populaire en interessante techniek in Natural Language Processing is Sentiment Analysis (Hirschberg & Manning, 2015). Sentiment Analysis is een soort van classificatie techniek waarbij een zin of document wordt geklassificeerd op basis van het onderliggende sentiment of gevoel van de tekst. Sentiment analysis kan opgedeeld worden in drie benaderingen: de lexicon gebaseerde benadering, de Machine Learning benadering, en de hybride benadering (Medhat et al., 2014). Bij de lexicon gebaseerde benadering werkt men met woordenboeken die aan tokens een bepaalde sentiment score geven, positief of negatief. Vele van deze tokens hebben een

score van nul (neutraal), maar positief geopinieerde woorden zoals “goed”, “wel”, “leuk”, krijgen een positieve waarde, en negatief geopinieerde woorden zoals “slecht”, “niet”, “saai”, krijgen een negatieve waarde. Op basis van deze scores van alle tokens in een zin of document wordt een algemene sentiment score berekend. Deze score kan zeggen of iets positiefs of negatiefs is, en ook hoe positief of negatief.

Bij de Machine Learning benadering wordt gebruik gemaakt van voorgetrainde Deep Learning modellen voor natuurlijke taal zoals BERT (Devlin et al., 2019). Deze voorgetrainde modellen worden verfijnd op sets van documenten die gelabeld zijn met “POSITIEF”, “NEGATIEF”, en soms ook “NEUTRAAL” (L. Zhang et al., 2018). Vele van deze verfijnde modellen in allerlei talen zijn te vinden op *Hugging Face* (n.d.). Als zo’n model wordt toegepast op een dataset van documenten, krijg je een classificatie “POSITIEF” of “NEGATIEF” en een zekerheid over deze classificatie. Deze zekerheid is echter geen score van hoe positief of negatief iets is.

Sentiment Analysis kan interessant zijn voor de LIM-methode om ook een idee te krijgen over hoe mensen staan tegenover de verkeerssituatie of andere onderwerpen waarover de onderzoeker meer wilt weten. Wij hebben dit hier niet kunnen toepassen om verschillende redenen. Ten eerste zijn de meeste NLP-modellen vooral rijk in de Engelse taal. Voor de Nederlandse taal zijn er op het moment van onze analyse niet veel Sentiment Analysis modellen vrij beschikbaar voor de Nederlandse taal en de modellen die beschikbaar waren, waren getraind op zeer specifieke datasets zoals boek beoordelingen, waardoor de accuraatheid op een ander soort dataset zoals internetenquêtes minder goed is door het verschil in taalgebruik. Dit zou opgelost kunnen worden door de teksten eerst te vertalen naar het Engels en dan een Engelstalig model op toe te passen. Verschillende auteurs hebben dit al uitgeprobeerd en vonden dat dit een competitieve aanpak is (Balahur & Turchi, 2016; Mohammad et al., 2016).

De tweede reden waarom wij Sentiment Analysis niet hebben kunnen toepassen ligt aan te vorm van de vraagstelling en de te verwachten antwoorden. Vraag 4 vroeg vooral naar de negatieve aspecten van wat er gevaarlijk is aan een bepaalde zone. Hierdoor kan je verwachten dat je vooral antwoorden zal krijgen met een negatief sentiment. Het kan misschien nog interessant zijn om een inzicht te krijgen hoe negatief een antwoord is om ergens een score te bepalen over hoe gevaarlijk het ergens is naast hoe vaak een gebied wordt aangeduid. Dit gaat echter niet met de Machine Learning benadering. Hiervoor zou dus met een lexicon gebaseerde benadering moeten gewerkt worden, maar opnieuw was er geen goed model te vinden voor de Nederlandse taal.

In de LIManalysis app ondervonden we bij de verkenning van de data dat er ook nog wat ruimte is voor verbeteringen in de vorm van extra info en filtermogelijkheden. Ten eerste zijn de vervoersmodi enkel gebruikt geweest als filter. Bij de voorbeelden in sectie 3.3 hadden wij vaak het idee dat het zeer interessant zou zijn om ook de frequenties van de vervoersmodi te zien waarop de andere filters van toepassing zijn. Bij sectie 3.3.2 zou het bijvoorbeeld interessant zijn om te zien hoe mensen zich naar de S8 verplaatsen. Bij sectie 3.3.3 kan je als onderzoeken je afvragen of het vooral fietsers of vooral voetgangers zijn die zich ergeren aan de snelheid. Dit is iets dat kan opgelost worden door bijvoorbeeld een histogram van de frequenties te presenteren op bijvoorbeeld de labels van de filter zelf of een nummer te zetten naast de labels die zeggen hoeveel er nog zijn na toepassing van de andere filters.

Ten tweede toont de LIManalysis tool bij de topics wel enkel de topics waarop de andere filters op toepassing zijn, maar je ziet niet welk topic het meest frequent voorkomt. Als je bijvoorbeeld op een plaats klikt, zoals nabij het tankstation, kan je het topic “snelheid” krijgen, maar je weet nu niet of dat topic zeer frequent daar voorkomt, of dat het maar één keer voorkomt. Het

kan bijvoorbeeld een uitschieter zijn van een andere aangeduide zone, zoals de oversteekplaats in het zuidwesten. Als dit maar één keer voorkomt, willen wij dit waarschijnlijk niet meenemen in onze interpretatie. Daarom is het interessant welk topic er het meest voorkomt en hoe vaak dit topic voorkomt. Dit kan door de topics de sorteren bij het voorstellen. De topic die het meest voorkomt als eerste zetten, enzovoort. Zo kan je sneller zien welke topics er belangrijker zijn. Ook zouden we kunnen bijzetten hoe vaak een topic voorkomt met onze andere filters. Zo kunnen we beter inschatten wat het relatieve belang is van een topic en welke topics we niet meer moeten meenemen in onze interpretatie.

### 4.3 Andere gebruiken

Natuurlijk is de LIM-methode niet gelimiteerd tot het bestuderen van fietsveiligheid. De LIM-methode kan gebruikt worden om eender welk ruimtelijk fenomeen te bestuderen aan de hand van de *knowledge of the crowd*. Het principe achter de LIM-methode zit hem erin om zoveel mogelijk informatie te behalen met zo weinig mogelijk moeite van de participanten. Hierin gaven wij een voorbeeld aan de hand van open vragen en een digitale markeerstift. Veel meer onderzoeken kunnen gedaan of versterkt worden met de LIM-methode.

Een eerste voorbeeld waarvoor de LIM-methode nog kan gebruikt worden is om de subjectieve drukte te beoordelen op festivals zoals Rock Werchter of de Gentse Feesten. Van de Weghe et al. (2013) hebben al verscheidene keren de bewegingen van mensen en de objectieve drukte opgemeten aan de hand van Bluetooth sensoren op Rock Werchter en de Gentse Feesten. Deze Bluetooth resultaten kunnen bijvoorbeeld vergeleken worden met een LIM studie waar gevraagd kan worden waar men het druk vindt op een bepaald moment en of ze de drukte op een bepaalde manier gevaarlijk vinden of niet. De Cock et al. (2023) hebben al een gelijkaardige studie gedaan op de Gentse Feesten

waarbij er continu op drie verschillende plaatsen in de festivalzone aan mensen werd gevraagd hoe druk ze het vonden op dat moment aan de hand van drie korte gesloten vragen, terwijl ze de eigenlijke drukte meten aan de hand van WiFi sensoren. Een extra voordeel dat LIM hierbij kan bieden is dat het onderzoek niet gelimiteerd moet zijn aan drie vooraf gekozen gebieden, en dat je met open vragen meer antwoorden kan krijgen over wat mensen van de drukte vinden.

Een ander voorbeeld waarvoor de LIM-methode kan gebruikt worden is in het geval van stadsplanning (Bugs, 2012; Jankowski et al., 2016). Stadsbesturen kunnen de LIM-methode gebruiken om meningen te verkrijgen van inwoners van de stad. Zo kunnen de inwoners of toeristen zonder al te veel moeite hun mening uiten over nieuwe locaties voor parken, publiek sanitair, drinkwatervoorzieningen, wandelpaden, etc. Nog een voorbeeld voor het gebruik van de LIM-methode is bij het openbaar vervoer (Bąkowska-Waldmann & Kaczmarek, 2021). Onderzoeksacties kunnen gevoerd worden op het openbaar vervoer om te weten te komen welke routes mensen nodig hebben, en welke routes er verbeterd kunnen worden.

## 5 BESLUIT

In dit onderzoek evalueren wij de Less Is More (LIM) methode. Less Is More houdt in dat je meer behaalt door minder te doen. In de zin van Public Participatory GIS (PPGIS) houdt dit in meer respondenten behalen en meer variatie in antwoorden kunnen behalen met minder moeite van de respondenten en van de onderzoekers. Wij proberen dit te behalen met een online vragenlijst dat kort en simpel is, en met open vragen dat voor veel vrijheid en variatie zorgt in de antwoorden. Met een digitale markeerstift kunnen respondenten makkelijk geografische informatie intekenen met veel vrijheid. De open vragen kunnen geanalyseerd worden met Natural Language processing technieken waardoor we er topics uithalen. De verkregen data kan

visueel geanalyseerd en geïnterpreteerd worden aan de hand van de LIManalysis tool. Deze toont een heatmap van de ingetekende geografische data welke kan gefilterd worden op basis van de verkregen topics uit de open vragen en andere gegevens die verzameld werden.

Uit de casestudie met als thema verkeersveiligheid is gebleken dat de LIM-methode een nuttige optie kan zijn voor exploratieve onderzoeken. We konden een redelijk aantal respondenten bereiken en interessante informatie uit de resultaten halen. Het clusteren van de open antwoorden om er topics uit te halen heeft wel nog veel ruimte voor verbetering. Anders kan de LIM-methode veel vragen beantwoorden, zoals "Waar vinden mensen het gevaarlijk als ze naar een bepaalde plaats rijden?", "Waar rijden weggebruikers te snel?" en "Waar is de beste plaats om verkeersspiegels te plaatsen?".

In vergelijking met uitgebreidere onderzoeken, zoals de Fietsbarometer (Storme et al., 2022), geeft de LIM-methode meer variatie in antwoorden en is het ruimtelijk accurater om de perceptie van de mensen vast te stellen. De LIM-methode is echter niet goed om een specifiek fenomeen te onderzoeken. Met een kwadranten analyse en een  $\chi^2$  Goodness-Of-Fit test ondervonden we dat LIM en de Fietsbarometer een andere ruimtelijke spreiding geven van antwoorden. De Fietsbarometer en de LIM-methode ondervonden dus een ander fenomeen, alhoewel de casestudie met de LIM-methode zich vooral focuste op gevaarlijke plaatsen terwijl de Fietsbarometer het toeliet om zowel veilig als gevaarlijk te beoordelen.

De LIM-methode toont wel mogelijkheden om andere academische studies te verrijken of als exploratieve studie. Ook voor beleid making is de LIM-methode een geschikte kandidaat om de mening van buurtbewoners en andere belanghebbenden. Men moet wel goed oppassen om de vragen goed en duidelijk te formuleren zodat er geen verwarring kan zijn over waarnaar

gevraagd wordt. Ook moet men goed nadenken over hoe je reclame maakt ervoor zodat mensen deze zeker invullen. Als de data goed verzameld raakt kan je deze makkelijk interactief inspecteren en interpreteren.

## REFERENTIELIJST

- Altuncu, M. T., Yaliraki, S. N., & Barahona, M. (2020). *Graph-based Topic Extraction from Vector Embeddings of Text Documents: Application to a Corpus of News Articles* (arXiv:2010.15067). arXiv. <http://arxiv.org/abs/2010.15067>
- Amine, A., Elberrichi, Z., Simonet, M., & Laboratory, E. (2010). Evaluation of Text Clustering Methods Using WordNet. *International Arab Journal of Information Technology*, 7(4).
- Bąkowska-Waldmann, E., & Kaczmarek, T. (2021). The Use of PPGIS: Towards Reaching a Meaningful Public Participation in Spatial Planning. *ISPRS International Journal of Geo-Information*, 10(9), Article 9. <https://doi.org/10.3390/ijgi10090581>
- Balahur, A., & Turchi, M. (2016). Multilingual Sentiment Analysis using Machine Translation? *Journal of Artificial Intelligence Research*, 55, 95-130.
- Blei, David, M., Andrew, Y. Ng., & Michael, I. J. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(Jan), 993-1022.
- Brown, G. (2012). Public Participation GIS (PPGIS) for Regional and Environmental Planning: Reflections on a Decade of Empirical Research. *URISA Journal*, 25(2), 7-18.
- Brown, G., Donovan, S., Pullar, D., Pocewicz, A., Toohey, R., & Ballesteros-Lopez, R. (2014). An empirical evaluation of workshop versus survey PPGIS methods. *Applied Geography*, 48, 42-51. <https://doi.org/10.1016/j.apgeog.2014.01.008>

- Brown, G., & Kyttä, M. (2014). Key issues and research priorities for public participation GIS (PPGIS): A synthesis based on empirical research. *Applied Geography*, 46, 122-136. <https://doi.org/10.1016/j.apgeog.2013.11.004>
- Brown, G., Montag, J. M., & Lyon, K. (2012). Public Participation GIS: A Method for Identifying Ecosystem Services. *Society & Natural Resources*, 25(7), 633-651. <https://doi.org/10.1080/08941920.2011.621511>
- Bugs, G. (2012). Assessment of Online PPGIS Study Cases in Urban Planning. In B. Murgante, O. Gervasi, S. Misra, N. Nedjah, A. M. A. C. Rocha, D. Taniar, & B. O. Apduhan (Eds.), *Computational Science and Its Applications - ICCSA 2012* (Vol. 7333, pp. 477-490). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-31125-3\\_36](https://doi.org/10.1007/978-3-642-31125-3_36)
- Bugs, G., Granell, C., Fonts, O., Huerta, J., & Painho, M. (2010). An assessment of Public Participation GIS and Web 2.0 technologies in urban planning practice in Canela, Brazil. *Cities*, 27(3), 172-181. <https://doi.org/10.1016/j.cities.2009.11.008>
- Ciganaitye, G., Mackutye - Varoneckienye, A., & Krilaviicius, T. (2014). Text Document Clustering. *International Journal of Design, Analysis and Tools for Integrated Circuits and Systems*, 5(2), 1-5.
- Cockton, G. (2012). *Usability Evaluation [2nd ed.]* (M. Soegaard & R. F. Dam, Eds.). <http://www.interaction-design.org>. [http://www.interaction-design.org/encyclopedia/usability\\_evaluation.html](http://www.interaction-design.org/encyclopedia/usability_evaluation.html)
- Connor Desai, S., & Reimers, S. (2019). Comparing the use of open and closed questions for Web-based measures of the continued-influence effect. *Behavior Research Methods*, 51(3), 1426-1440. <https://doi.org/10.3758/s13428-018-1066-z>

- Corbett, J., & Cochrane, L. (2020). Geospatial Web, Participatory. In *International Encyclopedia of Human Geography* (pp. 131-136). Elsevier. <https://doi.org/10.1016/b978-0-08-102295-5.10604-3>
- Czepkiewicz, M., Jankowski, P., & Zwoliński, Z. (2018). Geo-Questionnaire: A Spatially Explicit Method for Eliciting Public Preferences, Behavioural Patterns, and Local Knowledge - An Overview. *Quaestiones Geographicae*, 37(3), 177-190. <https://doi.org/10.2478/quageo-2018-0033>
- De Cock, L., Vandeviver, C., Verstockt, S., Treille De Grandsaigne, G., De Doncker, P., Huang, H., & Van de Weghe, N. (2023). *Comparing Crowding Perception and Sensor Counts at the Ghent Festivities*. <http://lib.ugent.be/catalog/pug01:01H40VS3QA9ATHRCJ9YN9RM6J0>
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391-407. [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<391::AID-ASI1>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9)
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. <https://doi.org/10.48550/arXiv.1810.04805>
- Eisenstein, J. (2019). *Introduction to Natural Language Processing*. MIT Press.
- Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211-221. <https://doi.org/10.1007/s10708-007-9111-y>
- Griffith, D. A., Amrhein, C. G., & Desloges, J. R. (1990). *Statistical Analysis for Geographers* (1st edition). Prentice Hall.

- Haklay, M., & Tobón, C. (2003). Usability evaluation and PPGIS: Towards a user-centred design approach. *International Journal of Geographical Information Science*, 17(6), 577-592.  
<https://doi.org/10.1080/1365881031000114107>
- Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014). Word cloud explorer: Text analytics based on word clouds. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 1833-1842.  
<https://doi.org/10.1109/HICSS.2014.231>
- Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, 349(6245), 261-266.  
<https://doi.org/10.1126/science.aaa8685>
- Hout, M. C., Papesh, M. H., & Goldinger, S. D. (2013). Multidimensional scaling. *WIREs Cognitive Science*, 4(1), 93-103.  
<https://doi.org/10.1002/wcs.1203>
- Hugging Face – The AI community building the future.* (n.d.). Retrieved July 3, 2023, from <https://huggingface.co/>
- ISO. (2009). 9241-210: 2010. Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems. *International Standardization Organization (ISO). Switzerland*.
- Jáñez-Martino, F., Alaiz-Rodríguez, R., González-Castro, V., Fidalgo, E., & Alegre, E. (2023). Classifying spam emails using agglomerative hierarchical clustering and a topic-based approach. *Applied Soft Computing*, 139, 110226. <https://doi.org/10.1016/j.asoc.2023.110226>
- Jankowski, P., Czepkiewicz, M., Młodkowski, M., & Zwoliński, Z. (2016). Geo-questionnaire: A Method and Tool for Public Preference Elicitation in

- Land Use Planning. *Transactions in GIS*, 20(6), 903-924.  
<https://doi.org/10.1111/tgis.12191>
- Lahitani, A. R., Permanasari, A. E., & Setiawan, N. A. (2016). Cosine similarity to determine similarity measure: Study case in online essay assessment. *2016 4th International Conference on Cyber and IT Service Management*, 1-6. <https://doi.org/10.1109/CITSM.2016.7577578>
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- Mohammad, S. M., Salameh, M., & Kiritchenko, S. (2016). How Translation Alters Sentiment. *Journal of Artificial Intelligence Research*, 55, 95-130. <https://doi.org/10.1613/jair.4787>
- Morante, R., & Blanco, E. (2021). Recent advances in processing negation. *Natural Language Engineering*, 27(2), 121-130. <https://doi.org/10.1017/S1351324920000534>
- Newman, G., Zimmerman, D., Crall, A., Laituri, M., Graham, J., & Stapel, L. (2010). User-friendly web mapping: Lessons from a citizen science website. *International Journal of Geographical Information Science*, 24(12), 1851-1869. <https://doi.org/10.1080/13658816.2010.490532>
- Peterfreund, E., & Gavish, M. (2021). Multidimensional scaling of noisy high dimensional data. *Applied and Computational Harmonic Analysis*, 51, 333-373. <https://doi.org/10.1016/j.acha.2020.11.006>
- Pocewicz, A., Nielsen-Pincus, M., Brown, G., & Schnitzer, R. (2012). An Evaluation of Internet Versus Paper-based Methods for Public Participation Geographic Information Systems (PPGIS). *Transactions in GIS*, 16(1), 39-53. <https://doi.org/10.1111/j.1467-9671.2011.01287.x>

- Popping, R. (2015). Analyzing Open-ended Questions by Means of Text Analysis Procedures. *BMS Bulletin of Sociological Methodology/ Bulletin de Methodologie Sociologique*, 128(1), 23-39. <https://doi.org/10.1177/0759106315597389>
- Ramírez Aranda, N., De Waegemaeker, J., Venhorst, V., Leendertse, W., Kerselaers, E., & Van de Weghe, N. (2021). Point, polygon, or marker? In search of the best geographic entity for mapping cultural ecosystem services using the online public participation geographic information systems tool, "My Green Place." *Cartography and Geographic Information Science*, 48(6), 491-511. <https://doi.org/10.1080/15230406.2021.1949392>
- Rogerson, P. A. (2015). *Statistical Methods for Geography: A Student's Guide* (4th ed.). SAGE Publications, Inc.
- Roghaniad, M. M., & Bohns, V. K. (2017). Ask in person: You're less persuasive than you think over email. *Journal of Experimental Social Psychology*, 69, 223-226. <https://doi.org/10.1016/j.jesp.2016.10.002>
- Ross, E. (2020, December 14). *Cosine Similarity is Euclidean Distance* [Blog]. Skeptric. <https://skeptric.com/cosine-is-euclidean/>
- Roth, R. E., Ross, K. S., & MacEachren, A. M. (2015). User-Centered Design for Interactive Maps: A Case Study in Crime Analysis. *ISPRS International Journal of Geo-Information*, 4(1), Article 1. <https://doi.org/10.3390/ijgi4010262>
- Sieber, R. (2006). Public Participation Geographic Information Systems: A Literature Review and Framework. In *Annals of the Association of American Geographers* (Vol. 96, Issue 3, pp. 491-507).

- Sieber, R. E., Robinson, P. J., Johnson, P. A., & Corbett, J. M. (2016). Doing Public Participation on the Geospatial Web. *Annals of the American Association of Geographers*, 106(5), 1030–1046. <https://doi.org/10.1080/24694452.2016.1191325>
- Sinha, G., Shahi, R., & Shankar, M. (2010). Human Computer Interaction. *2010 3rd International Conference on Emerging Trends in Engineering and Technology*, 1–4. <https://doi.org/10.1109/ICETET.2010.85>
- Smyth, J. D., Dillman, D. A., Christian, L. M., & McBride, M. (2009). Open-ended questions in web surveys. *Public Opinion Quarterly*, 73(2), 325–337. <https://doi.org/10.1093/poq/nfp029>
- Steinbach, M., Karypis, G., & Kumar, V. (2000). *A Comparison of Document Clustering Techniques* (Technical Report 00-034). Department of Computer Science and Engineering; University of Minnesota.
- Storme, T., Benoit, S., Van de Weghe, N., Mertens, L., Van Dyck, D., Brondeel, R., Witlox, F., Zwartjes, L., & Cardon, G. (2022). Citizen science and the potential for mobility policy – Introducing the Bike Barometer. *Case Studies on Transport Policy*. <https://doi.org/10.1016/j.cstp.2022.05.013>
- Sun, L., & Korhonen, A. (2011). Hierarchical Verb Clustering Using Graph Factorization. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*.
- Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.
- Thomas, R. W. (1977). *An introduction to quadrat analysis*. Geo Abstracts Ltd.
- Van de Weghe, N., Bellens, R., De Jaeger, T., Gautama, S., Huybrechts, R. Ug. 002005072145 802001130488, Meier, B., Versichele, M., Zlatanova, S., Peters, R., Dilo, A., & Scholten, H. (2013). *Towards an integrated crowd management platform*. <http://lib.ugent.be/catalog/pug01:5770937>

- Van Loon, A. (2022). *Three Families of Automated Text Analysis*.  
<https://doi.org/10.31235/osf.io/htnej>
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17, 261-272. <https://doi.org/10.1038/s41592-019-0686-2>
- Wang, Y.-X., & Zhang, Y.-J. (2013). Nonnegative Matrix Factorization: A Comprehensive Review. *IEEE Transactions on Knowledge and Data Engineering*, 25(6), 1336-1353. <https://doi.org/10.1109/TKDE.2012.51>
- Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301), 236-244. <https://doi.org/10.1080/01621459.1963.10500845>
- Wilson, A., Tewdwr-Jones, M., & Comber, R. (2019). Urban planning, public participation and digital technology: App development as a method of generating citizen involvement in local planning processes. *Environment and Planning B: Urban Analytics and City Science*, 46(2), 286-302. <https://doi.org/10.1177/2399808317712515>
- Yan, Y., Feng, C. C., Huang, W., Fan, H., Wang, Y. C., & Zipf, A. (2020). Volunteered geographic information research in the first decade: A narrative review of selected journal articles in GIScience. *International Journal of Geographical Information Science*, 34(9), 1765-1791. <https://doi.org/10.1080/13658816.2020.1730848>

- Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery*, 8(4), e1253.  
<https://doi.org/10.1002/widm.1253>
- Zhang, S. (2019). Public participation in the Geoweb era: Defining a typology for geo-participation in local governments. *Cities*, 85, 38-50.  
<https://doi.org/10.1016/j.cities.2018.12.004>
- Züll, C. (2016). Open-Ended Questions (Version 2.0). In *GESIS Survey Guidelines*. GESIS - Leibniz Institute for the Social Sciences.  
[https://doi.org/10.15465/gesis-sg\\_en\\_002](https://doi.org/10.15465/gesis-sg_en_002)