



Onderzoek; representativiteit

Één van de ons meest gestelde vragen luidt: Hoe representatief is de (fiets/wandel/hardloop) appdata? Oftewel, in hoeverre komen de intensiteiten van gebruik, resulterend uit visualisaties van appdata, overeen met de werkelijke gebruiksintensiteit?

In de laatste jaren hebben wij gebruik gemaakt van zowel Strava, Runkeeper als Endomondo. Endomondo is echter de grootste en best toegankelijke dataset en hebben we daarom het vaakst ingezet bij onderzoeken.

Uiteraard bevat deze dataset slechts een klein aandeel van de verplaatsingen in een gegeven regio. De absolute aantallen zullen daarom nooit een-op-een overeenkomen met de werkelijkheid. Maar dit is niet relevant, de vraag is in hoeverre activiteiten van appdata in vergelijkbare verhoudingen gebruik maken van paden en gebieden. Oftewel; horen de paden die in werkelijkheid het meest intensief gebruikt worden ook in de app-dataset tot de meest gebruikte paden? En als een gegeven pad in werkelijkheid

drie keer zo intensief wordt gebruikt als een ander pad, is een vergelijkbare intensiteitsverdeling dan ook in de appdata herkenbaar?

Er zijn, voor zover ons bekend, twee manieren om hier onderzoek naar te doen.

1. Door de kenmerken van de gebruikers en hun gedrag te onderzoeken.
2. Door de appdata te vergelijken met lokale (fiets)tellingen, waarin werkelijke aantallen passanten gedurende een bepaalde periode geteld zijn.

We hebben hier een aantal onderzoeken naar gedaan de afgelopen jaren, waarvan we de belangrijkste hier uitleggen.

1. Kenmerken van de gebruikers onderzoeken

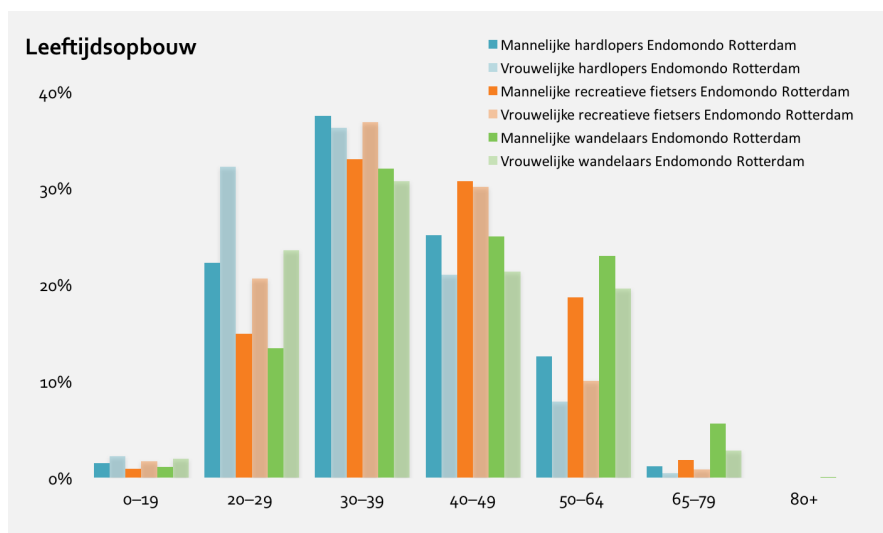
Kenmerken vanuit de Endomondo dataset zelf

In onderzoeken die we voor Den Haag en Rotterdam hebben gedaan, zijn de kenmerken van gebruikers van de app Endomondo vrij grondig uiteengelegd. Dat bleek in de twee steden weinig te verschillen. Voor de meest complete inzicht verwijzen we naar het rapport 'Routing Rotterdam', te downloaden via:

<https://drive.google.com/open?id=19RoZed767HxuQjRBGdK6xoADz15WE836>

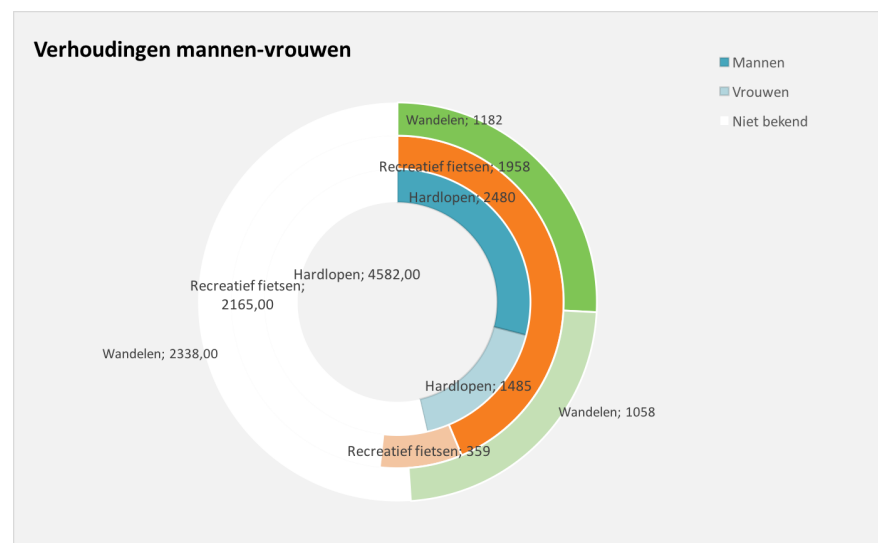
We laten hier enkele van de belangrijkste conclusies zien voor zowel fietsers, wandelaars als hardlopers.

Leeftijden Endomondo gebruikers



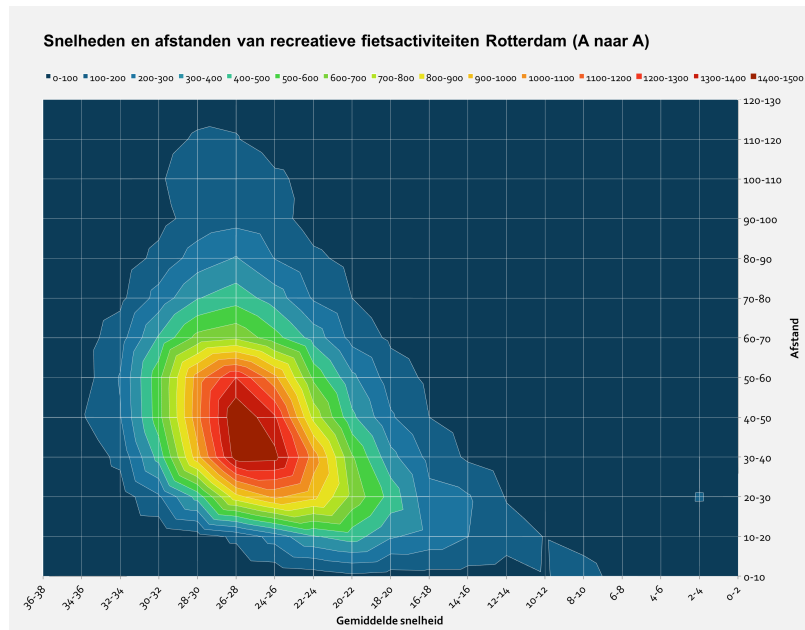
Anders dan wat soms gedacht wordt, blijkt Endomondo niet alleen gebruikt door 'jonge mensen'. Tussen de 20 en 65 jaar zit een redelijke spreiding. Juist jonger dan 20 jaar is ondervertegenwoordigd. Zeker onder wandelaars is het gebruik ook nog sterk tussen 50 en 64 jaar. 65 plus is wel ondervertegenwoordigd, maar deze groep is in het algemeen minder sportief.

Geslacht Endomondo gebruikers

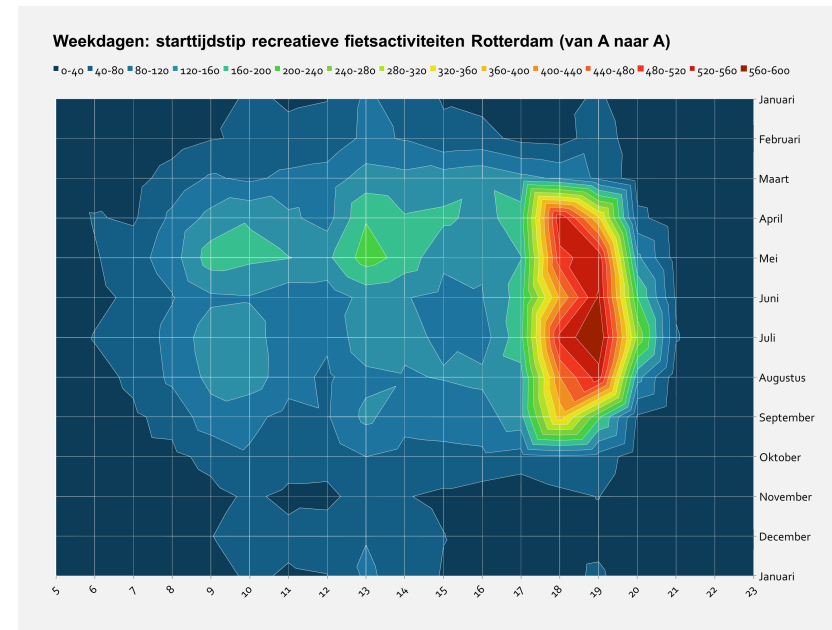


Ongeveer 50% van de Endomondo gebruikers heeft het geslacht aangegeven. Bij wandelen is deze verdeling behoorlijk gelijk. Bij hardlopen zijn er iets meer mannen onder de gebruikers. Bij recreatief fietsen hebben de mannen de overhand. Echter, dit is in gelijke mate het geval in de wielersport. In de volgende afbeelding is zichtbaar dat de recreatieve fietser vooral uit wielrenners bestaan.

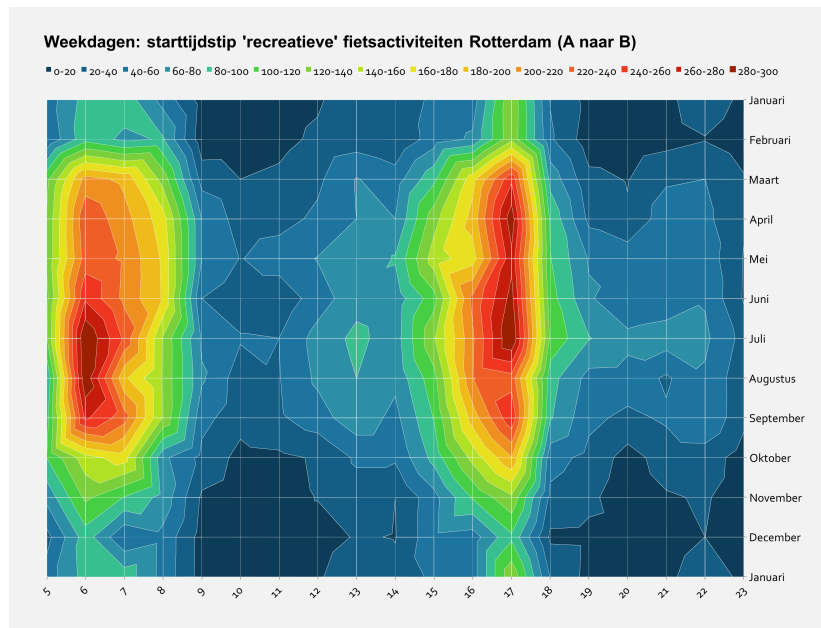
Fietsers, recreatief en utilitair



Recreatieve fietsers die de app Endomondo gebruiken blijken vooral een sportief karakter te hebben. De afstanden van hun ritten zijn vaak 30-60 kilometer en 24-30 kilometer per uur gemiddeld.

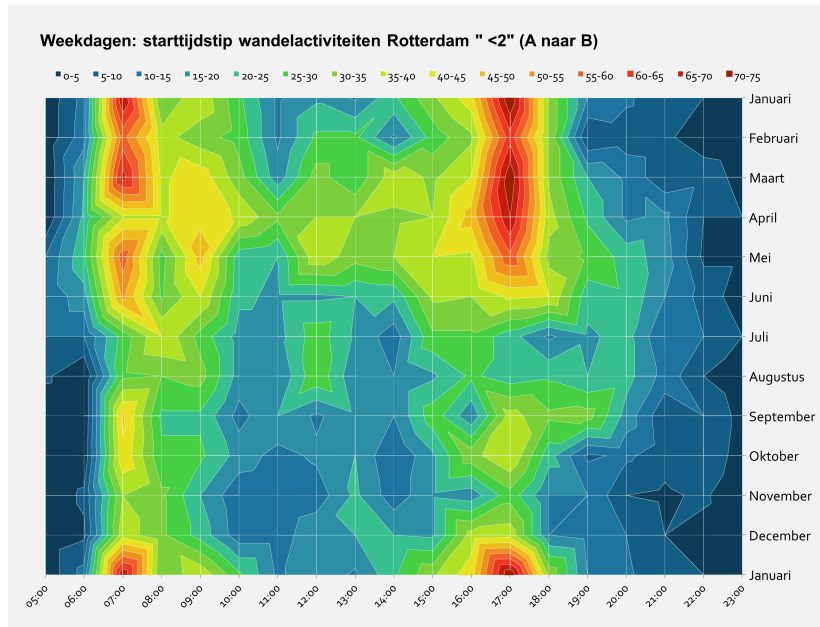


Doordeweeks wordt er vooral na het werk gefietst, een typisch kenmerk voor sportieve fietsers.

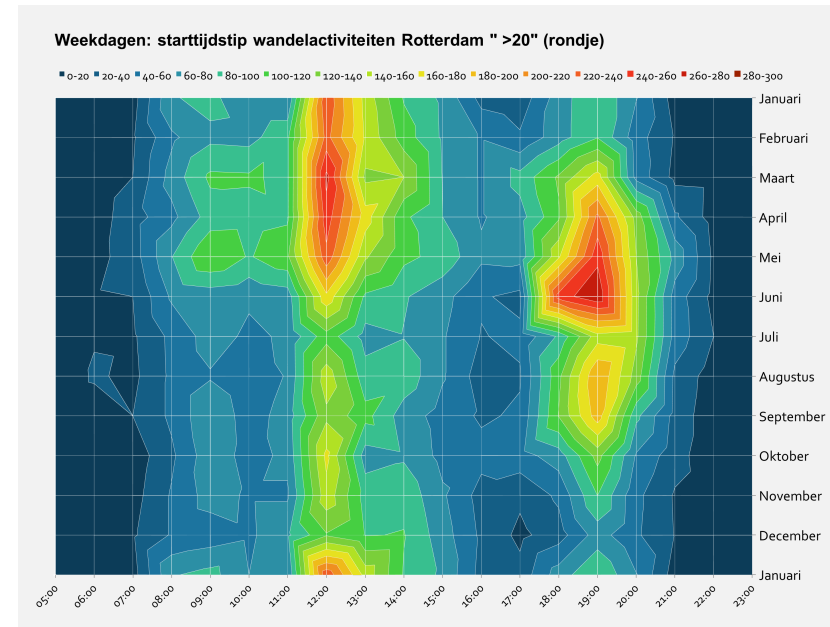


Er blijken echter ook fietsritten als recreatief te worden opgeslagen, die in werkelijkheid utilitair zijn. 'recreatieve fietsritten' waarbij de afstand tussen het begin en eindpunt van de route relatief groot was, vertrokken doordeweeks bijna allemaal tussen 6:00 en 8:00 uur of 16:00 en 18:00 uur. Deze ritten zijn in feite dus utilitair, en worden bij de dataset 'utilitair fietsverkeer' gevoegd.

Wandelaars



Anders dan bij fietsen worden wandelingen in de app Endomondo niet onderscheiden als zijnde 'recreatief wandelen' of 'utilitair wandelen'. Echter, door goed naar de kenmerken van wandelingen te kijken, is hier toch een onderscheid in te maken. Zo is bijvoorbeeld te herkennen dat wandelingen waarbij het begin- en eindpunt ver uit elkaar liggen (de afstand van de wandeling is minder dan tweemaal zo groot als de hemelsbrede afstand tussen begin- en eindpunt) doordeweeks vrijwel uitsluitend om 7:00 - 8:00 of 16:00-18:00 starten. Dit zijn dus herkenbare woon-werkwandelingen. Deze wandelingen zijn dan ook veelal niet meer dan 0,5-2 kilometer lang.

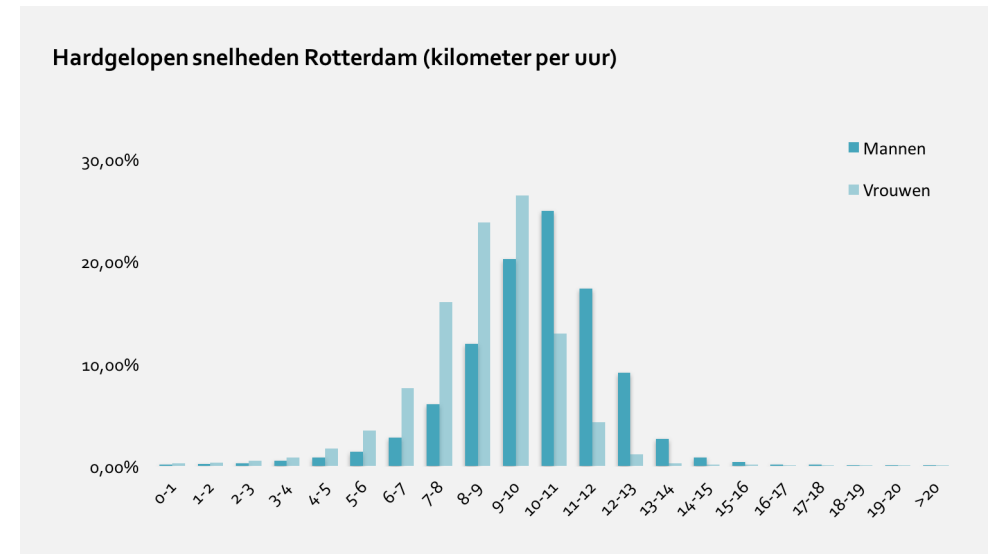
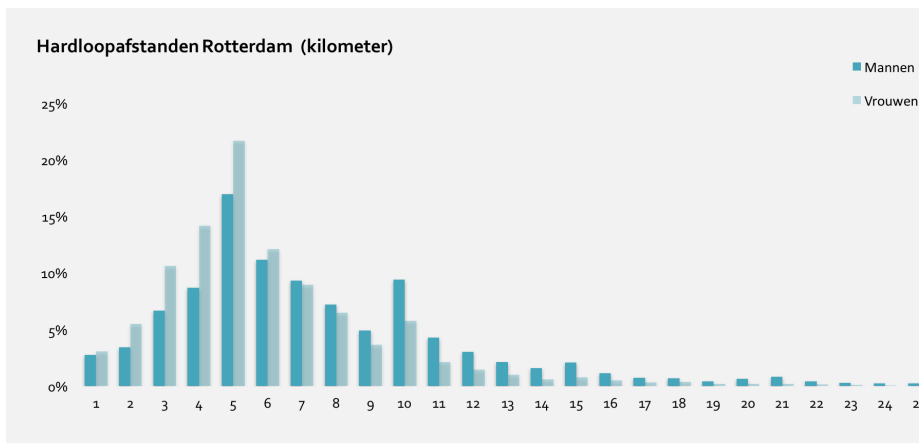


Omgekeerd kunnen we zien dat indien begin- en eindpunt wel dicht bij elkaar ligt (er wordt dus een rondje gelopen), de starttijden doordeweeks heel anders zijn. Ze beginnen of tussen 12:00 en 14:00 (lunchwandelingen) of na 18:00 (avondwandelingen). Deze wandelingen hebben vaak een afstand van 3 tot 10 kilometer.

Op basis van deze kenmerken, kunnen recreatieve en utilitaire wandelingen, behoorlijk goed van elkaar gescheiden worden.

Hardlopers

De 'hardloper' is al een veel specifiekere groep dan de 'fietser' of de 'wandelaar'. Echter zijn er wel hardlopers in allerlei soorten en maten. Een onderscheid dat ruimtelijk gezien belangrijk is, betreft de mate van 'gevorderdheid'. Gevorderde lopers, lopen zowel verder als sneller. Hun ruimtelijk bereik is groter, wat grote consequenties heeft op ruimtegebruik. We zien dat er bij de Endomondo gebruiker een behoorlijke spreiding is. Activiteiten zijn vaak 3-10 kilometer lang en 8-13 km/uur snel. De verspreiding binnen deze range is behoorlijk gelijk. Op basis van eigen inzicht lijkt dit een behoorlijk gemiddelde, representatieve groep hardlopers. Het is in elk geval zeker niet zo dat de app Endomondo alleen door 'fanatieke' hardlopers wordt gebruikt.



Kenmerken hardlopende appgebruikers op basis van wetenschappelijke literatuur

Een aantal (wetenschappelijke) instellingen heeft onderzoek gedaan naar gebruikers van activity tracking apps. Dit richtte zich met name op hardlopers. Er wordt hierin gekeken naar de loopmotivaties van appgebruikers en de motivaties om een app te gebruiken. Ook worden persoonlijke en demografische kenmerken van niet- en wel-appgebruikers vergeleken.

De TU Eindhoven heeft hier onderzoek naar gedaan in (*who uses running apps and sports watches? 2017*):

<https://pure.tue.nl/ws/portalfiles/portal/74875309>

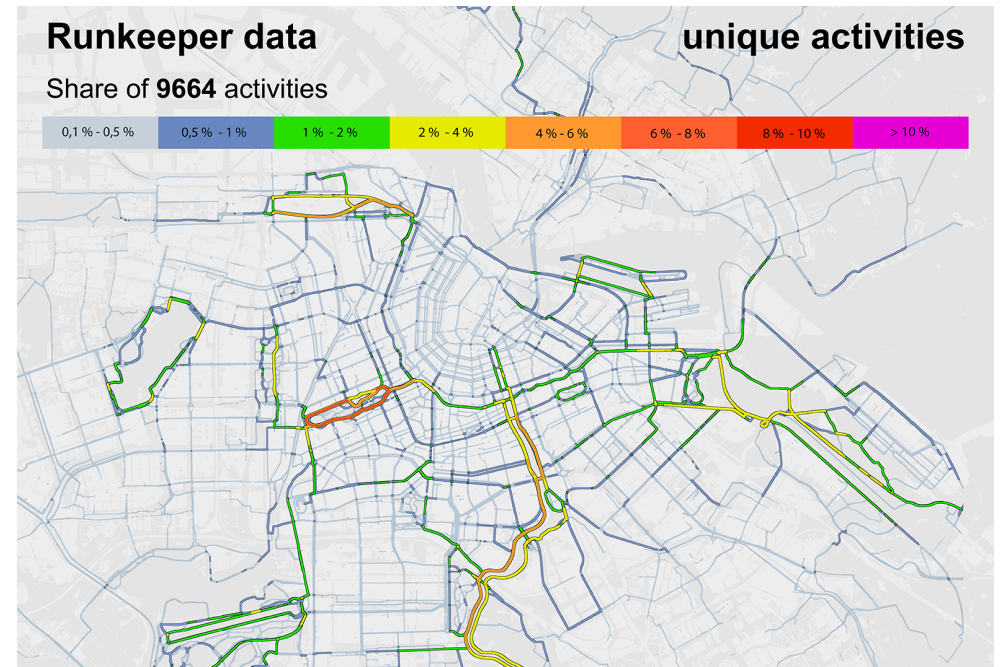
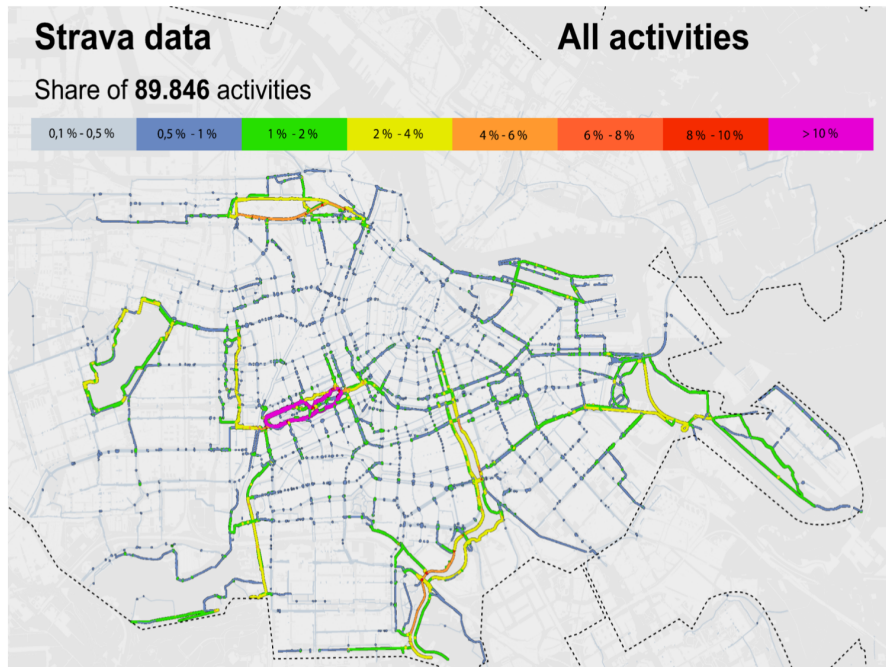
Daarin wordt bijvoorbeeld aangegeven dat apps minder worden gebruikt door oudere lopers. In relatieve zin zal dit zo zijn, we zien echter ook dat de groep 50-65 in de Endomondo data behoorlijk aanwezig is.

Ook is er (door de HVA) onderzoek gedaan naar appgebruik onder deelnemers van de Dam tot Damloop (*App use, physical activity and healthy lifestyle: a cross sectional study, 2015*). Hieruit blijken appgebruikers gemiddeld gezien relatief minder vaak te trainen (in voorbereiding op de Dam tot Damloop) dan niet-appgebruikers. Ze hebben ook iets vaker overgewicht. Daarmee lijkt de appgebruiker gemiddeld gezien iets minder 'gevoerd'. Echter, dit onderzoek richt zich sowieso op deelnemers van een loopwedstrijd.

De onderzoeken geven diverse nuanceverschillen aan tussen niet- en wel-appgebruikers. Er ontstaat echter geen beeld van bepaalde type hardlopers die niet of nauwelijks voorkomen in de groep 'appgebruikers'.

Vergelijking ruimtegebruik verschillende hardlooppapps

Een laatste manier om uitspraken te kunnen doen over representativiteit van hardloopdata, is het onderling vergelijken van ruimtegebruik van gebruikers van verschillende apps. In 2015 deden we dit in Amsterdam, met zowel Strava data (89.000 activiteiten) en Runkeeper data (9600 activiteiten). De beelden van ruimtegebruik waren zeer gelijkend, tot op zeer gedetailleerd niveau. De gebruikers van Runkeeper en Endomondo verschilden hier ook weinig in de afstanden die ze veel liepen. Meer data leidt dus niet perse tot een beter beeld, indien de belangrijkste kenmerken van gebruikers redelijk gelijk zijn. Loopafstand is dan ook de meest bepalende factor voor routekeuzes en keuzes in loopbestemmingen.



2. Door de appdata te vergelijken met lokale (fiets)tellingen

Lokale tellingen worden vooral gedaan op fietspaden, om aantallen daadwerkelijke fietspassages te meten. Tellingen van wandelpassages of hardlooppassages zijn er niet of nauwelijks.

Bij fietstellingen doet zich echter een 'probleem' voor. In de appdata zijn recreatieve ritten oververtegenwoordigd: appgebruik ligt onder recreatieve fietsers veel hoger dan onder functionele fietsers. Deze kunnen we in de Endomondo data van elkaar scheiden; als we een vergelijking willen maken is het nodig om functionele en recreatieve activiteiten apart van elkaar te vergelijken. Echter, bij de fietstellingen wordt doorgaans geen onderscheid gemaakt tussen functionele en recreatieve fietsers.

Vlaams Brabant

Daar is zover wij weten tenminste één uitzondering op, in de Provincie Vlaams-Brabant. Hier zijn (door de Directie ruimte dienst mobiliteit) op 13 locaties fietstellingen gedaan waarin ook gekenmerkt werd of het om een recreatieve of functionele fietser ging.

De studie is te vinden via: <http://docplayer.nl/4722048-Eindrapport-fietstellingen-kanaalroute-zenne-provincie-vlaams-brabant-directie-ruimte-dienst-mobiliteit.html>

Op de meetpunten werden (geordend van hoog naar laag daggemiddelde) onderstaande scores behaald, uitgedrukt in gemiddelde per dag.

Meetpunt	Gem_alle	recreatief	functioneel
BZ02R	413	153	260
BZ01R	363	138	225
BK07R	343	196	147
BK05R	261	131	131
BK06L	255	156	99
BK07L	181	114	67
BK02R	170	82	88
BK01L	165	64	101
BZ03R	194	99	95
BK00R	109	39	70
BK05L	173	109	64
BK06R	115	61	54
BK01R	88	26	62
SOM	2830	1366	1464

Op dezelfde telpunten hebben we uitgezocht hoeveel activiteiten er binnen de app-dataset passeerden.

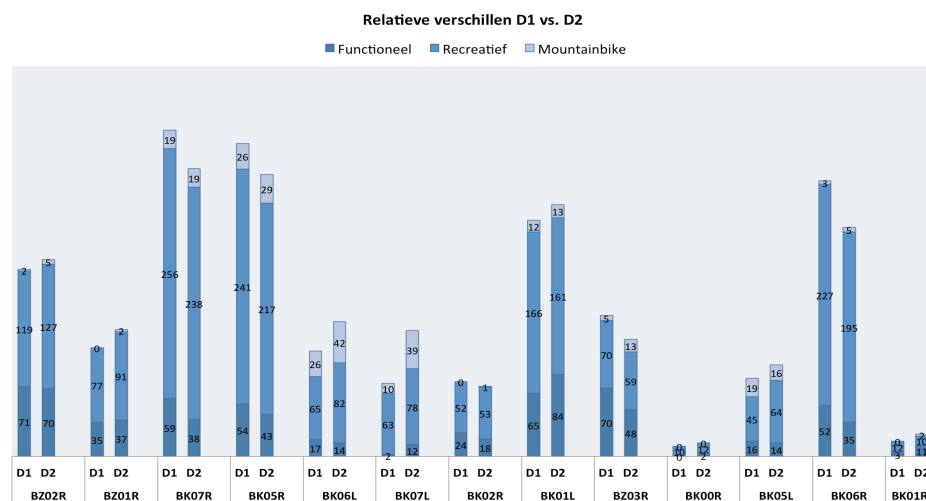
Dataset Endomondo

In dit onderzoek is data van de app 'Endomondo' vergeleken met de fietstelling. Hierbij was in eerste instantie de vraag welke tijdsperiode het beste genomen kan worden in de app-dataset. Het Vlaamse telonderzoek vond plaats in de maanden mei en juni van 2014. In het ideale onderzoek zou dan ook alleen activiteiten gedaan in mei of juni 2014 te tellen binnen de appdata,. Hier doet zich echter een mogelijk probleem voor; in een dergelijke korte periode is het aantal activiteiten binnen de appdata zeer beperkt (maximaal enkele tientallen ritten per telpunt). De andere mogelijkheid is dat we binnen de appdata, een wat grotere tijdsperiode nemen; bijvoorbeeld alle data van 2014 mee te nemen. Om te weten in hoeverre dit überhaupt uitmaakt, zou het goed zijn om de hoeveelheid activiteiten op de diverse meetpunten, uiteen te zetten voor twee tijdspannes: april 2014- juli2014 (D1). Dit is evenals de Vlaamse telling, de voorjaar/zomermaanden, weliswaar aan beide kanten een maand uitgebreid. En, geheel 2014 (D2).

Appdata

D1 (april-juli 2014)			D2 (heel 2014)		
functioneel	Sportief/ recreatief	mtb	functioneel	Sportief/ recreatief	mtb
41	68	1	71	127	5
20	44	0	37	91	2
34	147	11	38	238	19
31	138	15	43	217	29
10	37	15	14	82	41
1	36	6	12	78	39
14	30	0	18	53	1
37	95	7	81	161	12
40	40	3	48	59	13
0	6	0	2	12	0
9	26	11	14	64	16
30	130	2	35	195	5
2	7	0	11	10	2
269	804	71	424	1387	184

Uiteraard zijn de aantallen in D2 (heel 2014) in absolute zin groter. Om te kunnen vergelijken of het gebruik per pad relatief hetzelfde is gebleven of is veranderd, moet daarom eerst een correctie toegepast worden op het geheel. In D1 worden daarom recreatieve activiteiten vermenigvuldigd met 0,57 (=totaal aantal activiteiten in D1/ totaal aantal activiteiten in D2=1144/1995=0,57). Daaruit volgt het volgende resultaat:



Resultaat: de verschillen zijn klein, dat betekent dat het weinig uit zou maken of we D1 of D2 zouden gebruiken voor de vergelijking met de Vlaamse telresultaten. We nemen in dit geval D2.

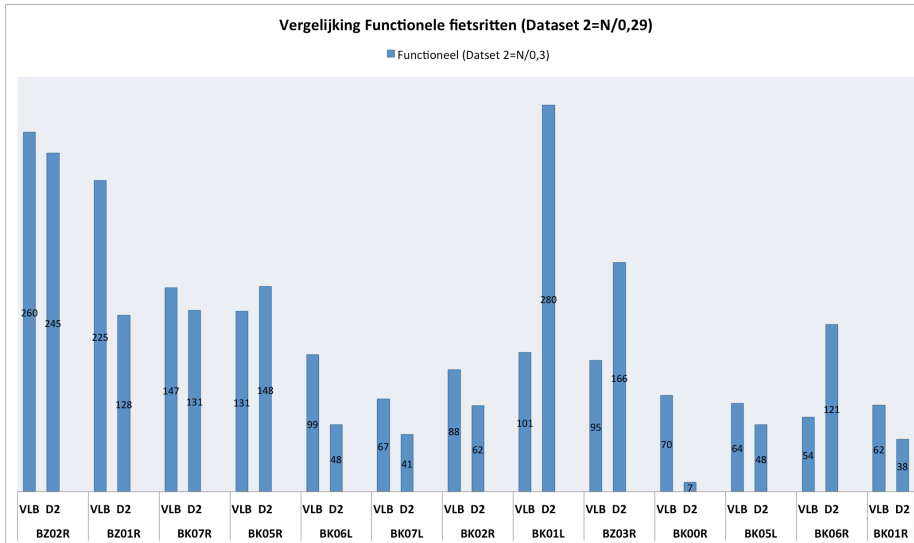
Vergelijking: functionele fietsritten

We beginnen met de vergelijking van de functionele fietsactiviteiten. Daarin zijn de cijfers als volgt:

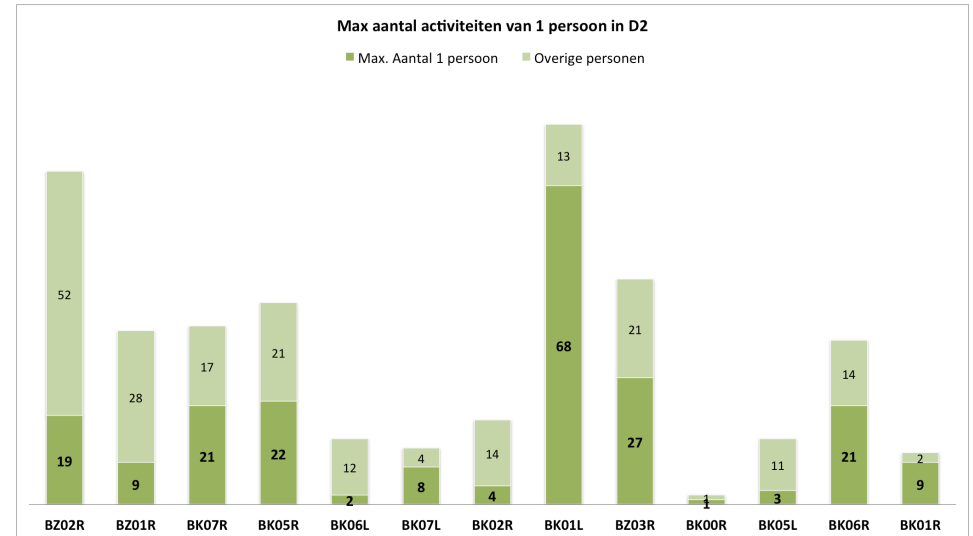
Meetpunt	VLB, telling	Endomondo, D2, aantal
BZ02R	260	71
BZ01R	225	37
BK07R	147	38
BK05R	131	43
BK06L	99	14
BK07L	67	12
BK02R	88	18
BK01L	101	81
BZ03R	95	48
BK00R	70	2
BK05L	64	14
BK06R	54	35
BK01R	62	11
SOM	1464	424

De eenheden zijn hierin wel verschillend, bij de VLB tellingen gaat het om daggemiddelden, in D2 om het totaal aantal passages in 2014. Dit is geen probleem (het gaat over een 'aantal passages per tijdseenheid'), maar om een vergelijking te kunnen maken dient wederom wel eerst een correctie toegepast te worden. De getallen van D2 worden in dit geval gedeeld door 0,29 (=totaal aantal activiteiten in D2/ totaal aantal activiteiten in VLB=1144/1995=0,29). De uitkomsten zijn dan als volgt:

Resultaat: Over het algemeen is het beeld dat de paden die binnen het VLB onderzoek intensief gebruikt werden, ook in D2 intensief gebruikt werden. Er zijn echter een aantal uitzonderingen: BKo1L, BZo3R en BKo6R. Hier scoort D2 significant hoger dan VLB.



In eerder onderzoek is al eens naar voren gekomen; dat een mogelijk 'gevaar' van appdata is, dat een enkel individu een grote invloed kan hebben op het algemene beeld. In D2 zijn de diverse punten tussen de 2 en 81 keer gepasseerd. Wanneer een enkel persoon dagelijks of wekelijks eenzelfde traject fietst (naar zijn werk bijvoorbeeld) en al deze activiteiten opneemt, vertegenwoordigt hij/zij al snel een groot aandeel van het aantal passages op een meetpunt. Dan kunnen vertekeningen snel ontstaan. We hebben daarom in D2 voor elk meetpunt uitgezocht, wat het maximaal aantal activiteiten van 1 persoon is:

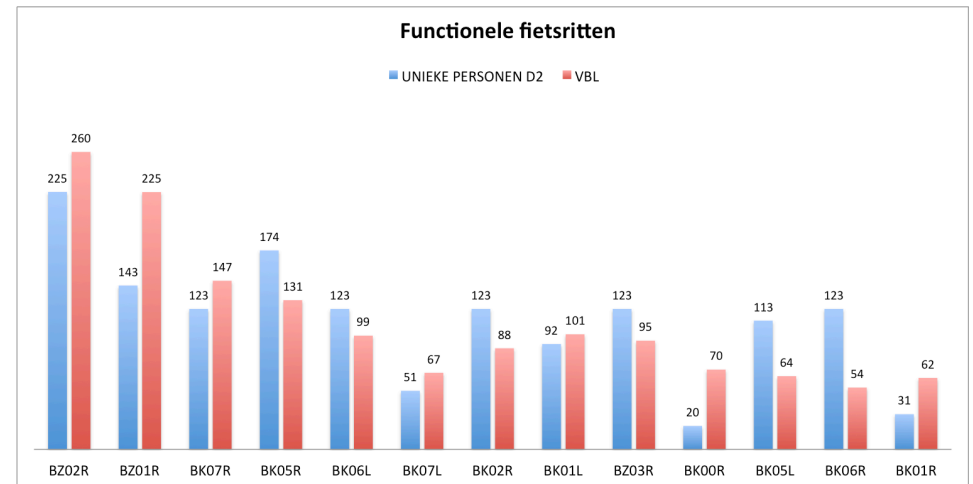


Dan wordt zichtbaar, dat bij zowel BKo1L, BZo3R als BKo6R, er sprake is van een hoog aantal activiteiten afkomstig van 1 individu. Bij BKo1L (waarbij het verschil tussen D2 en VLB het grootst was), is ook deze verhouding het hoogste, 68 van de 81 activiteiten waren van dezelfde persoon.

Daaruit volgt de conclusie; dat de representativiteit sterk verbeterd kan worden als er beperkingen gesteld worden aan het aantal activiteiten dat 1 persoon mag vertegenwoordigen binnen een dataset. Het is daarmee interessant om te bekijken wat de resultaten zijn, als we het aantal passages in VLB op de meetpunten, vergelijken met het aantal unieke individuen die in D2 op dezelfde meetpunten passeerden:

Meetpunt	D2: UNIEKE PERSONEN	VBL FUNCTIONEEL
BZ02R	22	260
BZ01R	14	225
BK07R	12	147
BK05R	17	131
BK06L	12	99
BK07L	5	67
BK02R	12	88
BK01L	9	101
BZ03R	12	95
BK00R	2	70
BK05L	11	64
BK06R	12	54
BK01R	3	62
	143	1464

Hier geldt wederom dat een correctie moet worden uitgevoerd, om de relatieve verdeling binnen de twee datasets over de diverse meetpunten te kunnen vergelijken. In dit geval $0,1 (= \text{totaal aantal unieke personen in D2} / \text{totaal aantal activiteiten in VLB functioneel} = 143 / 1464 = 0,1)$. Het resultaat is dan als volgt.



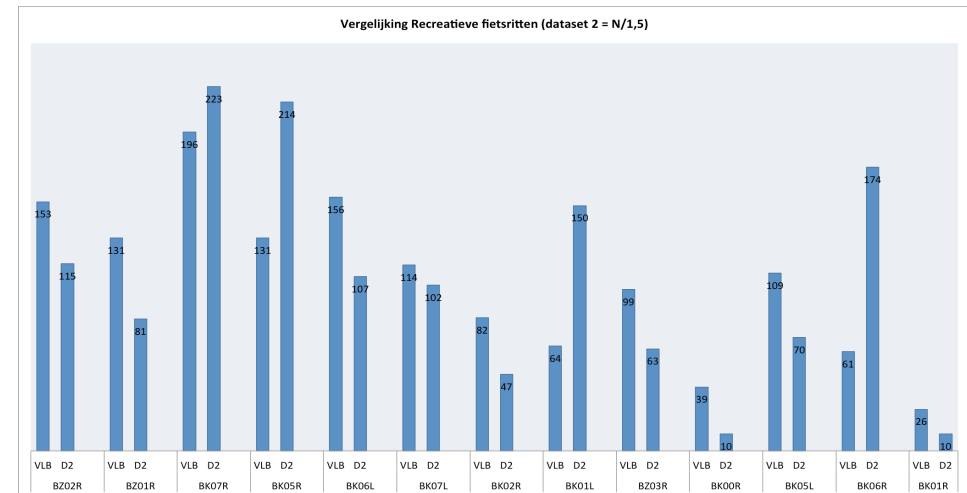
De gelijkenissen op de diverse meetpunten zijn wederom groot, maar nu zien we ook dat de uitschieters sterk gereduceerd zijn. De verschillen zijn nog het grootst in de laatste 4 meetpunten (BK00R, BK05L, BK06R, BK01R), maar hier is het aantal unieke individuen in D2 sowieso erg laag.

Vergelijking: recreatieve fietsritten

Nu volgt uiteraard de vraag, hoe zit dit bij recreatieve fietsactiviteiten? Volgens dezelfde methodiek:

Meetpunt	VBL, recreatief	Endomondo recreatief (incl mountainbiken)
BZ02R	153	132
BZ01R	138	93
BK07R	196	257
BK05R	131	246
BK06L	156	123
BK07L	114	117
BK02R	82	54
BK01L	64	173
BZ03R	99	72
BK00R	39	12
BK05L	109	80
BK06R	61	200
BK01R	26	12
SOM	1366	1571

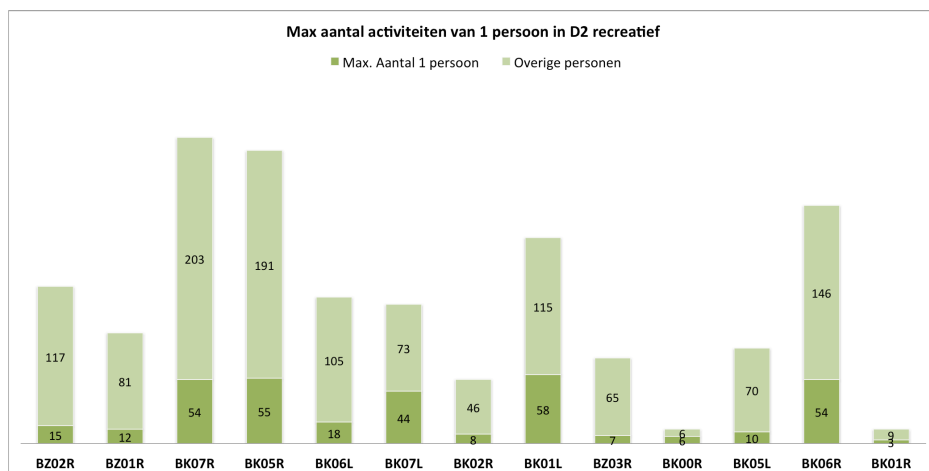
Correctie is in dit geval 1,15 (=1571/1366).



Wederom zijn de meeste meetpunten behoorlijk gelijk, en zijn er enkele uitschieters.

In hoeverre wordt deze vertekening bij recreatieve ritten veroorzaakt door de oververtegenwoordiging van een enkel individu?

Ook hier zien we weer eenzelfde patroon: daar waar D2 relatief meer activiteit op een meetpunt heeft dan VBL (BK05R, BK01L, BK06R), is sprake van een relatief grote vertegenwoordiging van een enkel persoon. Niet verrassend vereffent dit effect als we het aantal passages in VBL op de meetpunten, vergelijken met het aantal unieke individuen die in D2 op dezelfde meetpunten passeerden:

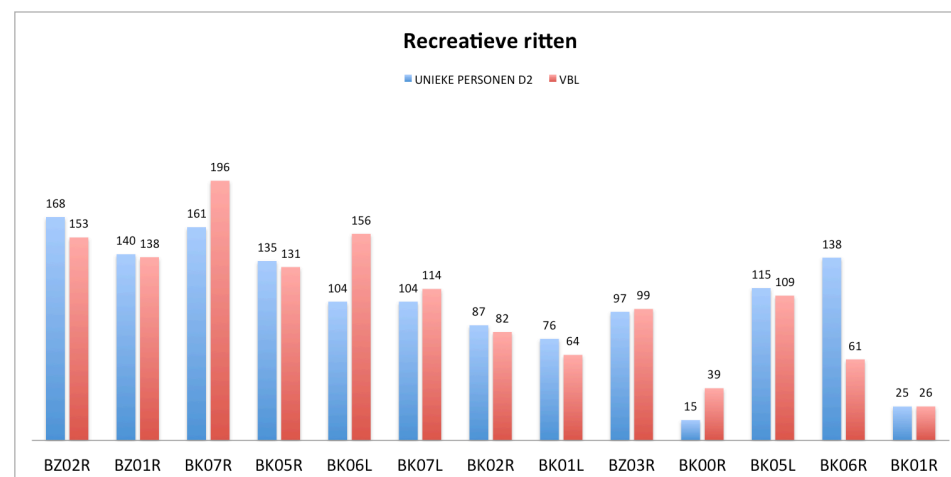


De correctie is in dit geval 2,55 (=1366/536).

De verschillen tussen VBL en D2 zijn op vrijwel alle meetpunten zeer klein, nog kleiner dan het geval was bij functionele fietsritten.

Hieruit volgt ook dat de verhoudingen recreatieve ritten/functionele ritten op de diverse meetpunten in D2, grote gelijkenissen heeft met VBL (wederom na het toepassen van een correctie op het geheel):

Meetpunt	2: UNIEKE PERSONEN	VBL
BZ02R	66	153
BZ01R	55	138
BK07R	63	196
BK05R	53	131
BK06L	41	156
BK07L	41	114
BK02R	34	82
BK01L	30	64
BZ03R	38	99
BK00R	6	39
BK05L	45	109
BK06R	54	61
BK01R	10	26
	536	1366



Fietstellingen Utrecht

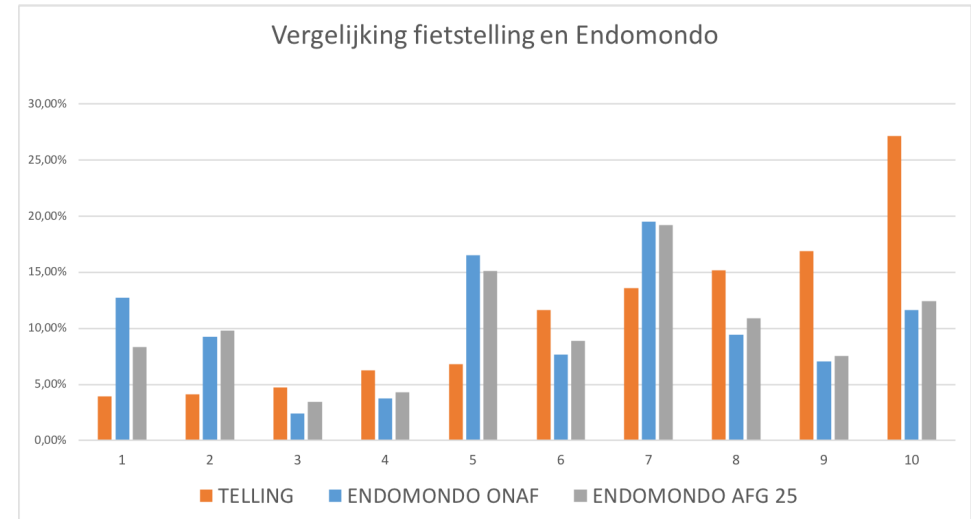
In Nederland zijn dergelijke tellingen, waarin recreatief en utilitair fietsverkeer gescheiden zijn, niet bekend. Echter, zeker binnenstedelijk kan aangenomen worden dat (zeker doordeweeks) het overgrote deel van het fietsverkeer utilitair is. We zouden dus wel een vergelijking kunnen maken tussen Nederlandse binnenstedelijke fietstellingen en de 'Endomondo utilitaire' fietsdata.

In Utrecht zijn de gegevens van fietstellingen openbaar en dus bruikbaar. Dit betreft 11 telpunten/tellocaties. Alleen het telpunt van 'de Vleutenseweg' hebben we niet meegenomen, omdat ons niet duidelijk is waar dit meetpunt ligt.

De Utrechtse tellingen zijn van 06-2015 t/m 12-2015. Bij Endomondo is wederom een wat ruimere tijdspanne genomen van 2014, 2015 en 2016, omdat anders de hoeveelheden data erg klein zijn. Overigens is wel bekeken of het veel uitmaakt welke tijdspanne genomen wordt, maar dit maakt ook hier niet veel uit. Wel is er rekening gehouden met wekdagen: we tellen alleen ritten op doordeweekse dagen in zowel de Endomondo data als de Utrechtse fietstellingen. Want, (en dat zal in het volgende hoofdstuk ook blijken) de Endomondo data lijkt vooral doordeweeks 'zuiver' utilitair te zijn.

We drukken ditmaal de hoeveelheid tellingen uit in 'aandeel van het totaal aantal tellingen', dat een telpunt vertegenwoordigt. Zo kunnen de tellingen en de Endomondo data nog makkelijker vergeleken worden.

De vergelijking in Vlaams-Brabant liet zien dat in de Endomondo data, enkele individuen een grote invloed kunnen hebben op het aantal passages van een plek/route. In Utrecht kijken we eerst in hoeverre dit uitmaakt:

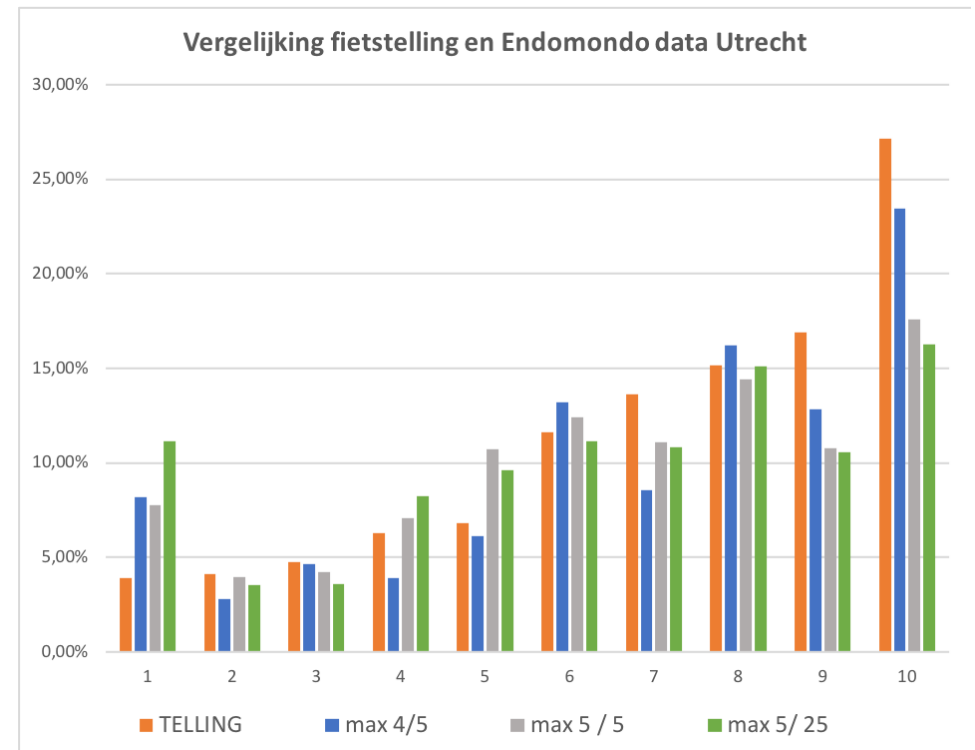


De oranje balken weergeven de aandelen van de Utrechtse tellingen. De blauwe balken weergeven de Endomondo aandelen, als we geen correctie toepassen op het aantal activiteiten van één individu. De grijze balken weergeven de Endomondo aandelen, als we wel een correctie toepassen op het aantal activiteiten van één individu. Het maximum is gesteld op 25 activiteiten van één persoon. Hoewel de grijze balken iets vaker in de buurt zitten van de oranje, is er geen sterk verband zichtbaar tussen de Utrechtse tellingen en de Endomondo tellingen. Ook als we de correctie 'strenger maken', bijvoorbeeld maximaal 10 of 5 activiteiten per individu, wordt dit niet aanzienlijk beter.

Daarmee lijkt de Endomondo data op het eerste oog niet bepaald representatief, zoals het in Vlaams Brabant wel lijkt. Echter, er is een belangrijk verschil tussen de vergelijking in Vlaams Brabant en Utrecht. De meetpunten in Vlaams Brabant zijn gelegen tussen dorpskernen en niet binnenstedelijk. Hier passeren logischerwijs relatief meer fietsritten die een langere afstand afleggen. We weten dat het aandeel appgebruikers ook groter is onder langere fietsritten, voor een hele korte fietsrit zijn mensen minder snel geneigd om een app aan te zetten (hun beweging te gaan meten). Het VBL onderzoek past daarom in schaal mogelijk beter bij de Endomondo/app-datasets.

Binnenstedelijk lijkt dit dus anders te liggen. Hier vormen korte fietsritten (naar verwachting) een groot aandeel van het fietsverkeer, en die zijn in onze datasets minder sterk vertegenwoordigd. En, bepaalde binnenstedelijke routes (en telpunten) zijn sterker onderdeel van lange, doorgaande routes dan andere. Dit kan sterk bepalen in welke mate ze door lange danwel korte afstandsfietsers worden aangedaan.

De locaties van de telpunten in Utrecht kunnen we echter niet veranderen, telpunten buiten de stad zijn er naar ons weten niet. Wat we echter wel kunnen doen, is in de Endomondo data, alleen 'korte' fietsritten meenemen in de telling. Van bijvoorbeeld maximaal 5 kilometer of zelfs korter. De meeste binnenstedelijke fietsritten zullen eerder 'enkele kilometers' lang zijn. Echter, slechts 14% van de Endomondo utilitaire fietsritten zijn korter dan 5 kilometer. De dataset waarmee vergeleken kan worden, is zeer beperkt qua formaat. De correctie voor 'maximaal aantal activiteiten per individu' zal naar verwachting ook strenger moeten zijn.



De vergelijking blijkt nu inderdaad een veel sterker verband te hebben. De Endomondo tellingen (blauw, grijs en groen) komen overall aanzienlijk dicht in de buurt van de Utrechtse tellingen (oranje). De groene balk (max 5 kilometer en max 25 activiteiten per individu), zit er meestal nog het verst naast. De representativiteit wordt sterker als het maximaal aantal activiteiten per individu 'streng' is (in dit geval 5). De blauwe balk komt het vaakst dicht in de buurt van oranje. Een maximum van 4 kilometer lijkt dus ook beter resultaat te geven dan een maximum van 5 kilometer.

Conclusies vergelijkingen fietstellingen en Endomondo fiets-data

Bij zowel functionele als recreatieve fietsritten uit de app-dataset, zijn er sterke gelijkenissen met de fietstellingen uit het Vlaamse onderzoek. Zeker bij recreatief fietsen is de gelijkenis tussen het VBL onderzoek en de app-data zeer groot. Het is te verwachten dat dit ook te maken heeft met het grotere aantal recreatieve activiteiten binnen de app-dataset.

De vergelijking in Utrecht laat zien hoe specifiek die representativiteit ligt. Een hele kleine dataset kan een behoorlijk representatief beeld geven indien de fietser specifiek gedefinieerd wordt. Ditzelfde was herkenbaar in de vergelijking van Strava en Runkeeper data bij hardlopers. De Endomondo utilitaire fietsdata zal meer de langere fietsbewegingen weergeven dan de korte. Echter, als je de korte fietsritten selecteert kun je ook deze al behoorlijk in beeld brengen, op basis van slechts hele kleine datasets.

Belangrijk is in elke geval dat er sterke vertekeningen ontstaan als een individu een groot aandeel vertegenwoordigt van het totaal aantal activiteiten op een meetpunt. Door dit aspect te corrigeren, worden deze uitschieters sterk teruggedrongen en ontstaan een waarheidsgetrouwer beeld.

Wat betekent dit voor het gebruik/buikbaarheid van de data?:

- Het geeft wat utilitair fietsverkeer vooral een beeld van langer, inter-stedelijk fietsverkeer.

- Het stellen van een maximum aan het aantal activiteiten dat een individu binnen een dataset kan hebben vergroot de waarheidsgetrouwheid, en is dus essentieel.
- Er kunnen afwijkingen zijn van enkele tientallen procenten. Dus; baseer geen beslissingen op een gedragspatroon waarin 10% verschil zit. Bijvoorbeeld; investeren in fietspad A i.p.v. B, omdat A in de data 10% meer gebruik toont dan fietspad B, is niet sterk. Echter, als fietspad A in de appdata 200% meer gebruikt wordt dan B, is het niet waarschijnlijk dat de paden in werkelijkheid evenveel gebruikt worden. Bij een dergelijke ordegrootte van verschillen kan het een sterkere basis vormen voor een beslissing.
- Datasets van dit formaat lenen zich niet zo zeer om kleine, korte termijn trends in gebruik te monitoren, zoals dagelijkse, wekelijkse of maandelijkse veranderingen in gebruik. Echter wel om jaargemiddeldes te visualiseren, of trends door jaren heen.

Deze vergelijking betreft natuurlijk (slechts) een bepaalde regio in België en één Nederlandse stad. Vergelijkingen op meer plekken zouden altijd wenselijk zijn.