

Tartu Ülikool
Loodus- ja täppisteaduste valdkond
Ökoloogia ja maateaduste instituut
Geograafia osakond

Bakalaureusetöö geograafias (12 EAP)

**GPS-andmete põhjal arvutatud tegevusruumide kujutamise meetodite
võrdlev analüüs**

Jürgen Öövel

Juhendaja: kaasprof Anto Aasa

Tartu 2024

Annotatsioon

GPS-andmete põhjal arvutatud tegevusruumide kujutamise meetodite võrdlev analüüs

Uue paradigma saabumine inimgeograafiasse tõi kaasa uute andmetüüpide ja analüüsimeetodite kasutuselevõtu. Samas on jäänud tähelepanuta nende meetodite endi olemuse uurimine. Selle töö eesmärgiks on uurida inimese tegevusruumi kujutamiseks kasutatavate meetodite loomuomast kallutatust. Selleks uuriti töö raames Kalamaja ja Priisle elanike GPS- andmete põhjal arvutatud ruumikujusid. Ruumikujude vaheliste erinevuste leidmiseks võrreldi erinevate ruumikujude pindalade mediaanväärtusi ning kasutati Duncani testi ja Kruskal-Wallise testi. Töö tulemustena selgub, et üks ruumikuju, puhvriga joon, näitab uuritud tunnuste puhul ülejäänud kolmest ruumikujust märkimisväärselt väiksemaid erinevusi võrdlusgruppide vahel. Töö panustab teadusmaailma arengusse teadmise, et kõik tegevusruumide kujutamiseks kasutatavad meetodid ei näita gruppide vahelisi erinevusi samas ulatuses, mis võimaldab järgmistel uuringutel teha teadlikumaid metoodika valikuid.

Märksõnad: GPS, tegevusruum, metoodika

CERCS kood: S230 – Sotsiaalne geograafia

Abstract

A comparative analysis of GPS-data based methods used for representing activity spaces

Arrival of a new, human mobility centric, paradigm brought with it the adoption of new data types and methods. However, the methods themselves have not been studied thoroughly. This thesis investigates the methods, that are used to describe our daily activity space, using GPS-data collected from the residents of Kalamaja and Priisle. Duncan test, Kruskal-Wallis test and comparing median values of the area sizes of different activity space methods was used to find differences between different methods. This study finds that of the four methods analysed only one, daily path area, consistently showed smaller differences between groups than the rest of the methods. This study is significant as it shows that some methods that are used for describing activity spaces do have inherit biases and this is something that must be considered when choosing the analysis methods.

Keywords: GPS, activity space, methodology

CERCS code: S230 – Social geography

Sisukord

Sissejuhatus	4
1. Teoreetiline ülevaade	6
1.1 Tegevusruumide uurimine	6
1.2 Mobiiliandmete kasutamine	6
1.3 Ruumikujude kasutamine	7
1.3.1 Ruumikujude arvutamine	9
1.4 Varasemad üliõpilastööd	10
2. Andmed ja töö käik	11
2.1 Andmete kogumine	11
2.1.1 Andmete kirjeldus	13
2.1.2 Rakendus MobilityLog	15
2.1.3 Andmekaitse	15
2.2 Andmetöötlus ja -analüüs	16
2.3 Tehisintellekti kasutamine	17
3. Tulemused	17
3.1 Korrelatsioonimaatriks	17
3.2 Mediaanväärtuste võrdlus	19
3.3 Tunnusepõhine võrdlemine	23
4. Arutelu	27
4.1 Tulemuste analüüs	27
4.2 Soovitused tulevasteks uurimusteks	27
5. Kokkuvõte	29
Summary	31
Tänuavaldused	33
Kasutatud kirjandus	34
Lisa 1. Mediaanväärtuste võrdlustabelid	39
Lihtlitsents	42

Sissejuhatus

Tänapäevane infoühiskond tugineb igapäevaelu toimimiseks üha enam suuremahuliste andmete kasutamisele. Suurandmed ning nende tõlgendamine mõjutab meie elu kõiki aspekte, alates sellest, milliseid reklaame me enda sotsiaalmeedias näeme kuni selleni, milliseid tänavaid pidi me töölt koju sõidame. Hoolimata kõigest kasulikust, mis suurandmete kasutamisega kaasneb on sellel siiski ka oma varjuküljed.

Kliimamuutuste valguses kerkib esile küsimus, milliste andmete säilitamine ja kasutamine on üldse vajalik, sest serverid, milles andmeid hoitakse tarbivad elektrienergiat. Rahvusvahelise Energiaagentuuri hinnangul moodustavad andmekeskused maailma elektrikasutusest juba 1-1,3% (IEA, i.a.). Sarnaselt andmete hoidmisele on energiakulukas ka nende andmete kasutamine. Andmete töötlemiseks vajalik energiahulk sõltub suuresti sellest, kui keerulisi arvutusi tehakse ning kui suure mahuga on töödeldavad andmed.

Lisaks keskkonnamuredele kaasnevad suurandmetega ka eetilised probleemid. Viimaste aastate tehisaru vaimustus on toonud teravalt esile nn suletud kasti probleemi. Tulenevalt tehisaruvõimekusest leida andmetest inimesele tavaliselt nähtamatuid seoseid ei suuda inimesed enam mõista, miks tehisaruvõime jõuab (Brožek *et al*, 2023).

Ülekantud tähenduses esineb seda probleemi ka geograafias ning loodusteadustes laiemalt. Suurandmeid kasutades on lihtne usaldada ning kasutada varem väljatöötatud meetodeid ning lahendusi, sest need tunduvad universaalsed. Siiski peitub keeruliste valmitite ning algoritmide taga oht, et meetodi rakendaja ei ole teadlik kasutatavasse meetodisse sisseehitatud kallutatusest, mistõttu võidakse tulemusi kogemata valesti tõlgendada. Seetõttu on oluline teada ning tunda erinevate meetodite mõju tulemustele, et takistada kallutatud tulemuste valede järeldusteni jõudmist.

Mobiilsusuuringutes on võtmetähtsusega inimeste ruumikasutuse kaardistamine. Ruumikasutuse mõistmine võimaldab analüüsida erinevaid aspekte inimese elust, alates ligipääsetavusest erinevatele teenustele ning lõpetades keskkonna- ja tervisemõjudega. Inimese tegevusruumi kirjeldamiseks ning kujutamiseks kasutatakse valdavalt erinevaid kujundeid – vähim ümbritsevat polügoni (VÜP, i.k. *minimum convex polygon*), vähim ümbritsev ristkülik (VÜR, i.k. *bounding box*), standardhälbe ellips (SHE, i.k. *standard deviation ellipsis*) ja puhvriga joon (i.k. *path area*). Varasemalt pole siiski uuritud, kas nendel kujutusmeetoditel esineb iseloomulik *bias* ehk kalduvus, mis moonutab saadavaid tulemusi.

Uurimistöö eesmärgiks on tuvastada, kas ja milline on erinevate inimese ööpäevast ruumikasutust kirjeldavate meetodite kallutatus erinevate sotsiaalsete tunnuste alusel moodustatud gruppide võrdlemisel. Lisaks uuritakse millised on kõige kuluefektiivsemad meetodid gruppide vaheliste erinevuste kirjeldamiseks. Eesmärkide täitmiseks on püstitatud järgnevad uurimisküsimused:

- Kas ja millised tegevusruumi kujutamise meetodid kujutavad kallutatult erinevusi erineva emakeele, elukoha ja haridustasemega gruppide tegevusruumide vahel?
- Millise tegevusruumi kujud on kõige kuluefektiivsemad ööpäevase ruumikasutuse erinevuste kujutamiseks?

1. Teoreetiline ülevaade

1.1 Tegevusruumide uurimine

Ruumi uurimine on ajalooliselt olnud geograafia, kui teadusvaldkonna põhieesmärk. Geograafia algusaegadel keskendus geograafia füüsilise ruumi uurimisele ning looduse kaardistamisele, peamiselt tegeleti maakaartide ja atlaste koostamisega. Inimese tegevusruumi uurimine ja kaardistamine on geograafias pigem uus uurimissuund, mis on erilise tähelepanu alla sattunud seoses ülemaailmse populatsioonikasvu ning linnastumise taustal.

Esmalt keskendus inimeste liikumise uurimine kindlate ruumipunktide põhiste meetoditele, kuid tänu ajageograafia teooria loomisele Rootsi geograafi, Torsten Hägerstrandi, poolt võeti kasutusele inimesekeskseid meetodeid (Neutens *et al*, 2007). Hägerstrand (1970) leidis, et lisaks inimtegevuse uurimisele geograafilises ruumis geograafiliste koordinaatidega tuleks uurida ka nende tegevuste ajalist järjestust ning korduvust ehk ka inimeste tegevuste ajalist dünaamikat.

Hägerstrandi teoorias (1970) on iga inimese liikumine piiratud kolme tüüpi piirangute poolt: võimekuse piirangud, ühildumise piirangud ja autoriteedi piirangud. Võimekuse piirangud on seotud inimese füüsiliste ja materiaalsete võimetega ruumis liikuda, näiteks auto olemasolu suurendab liikumisvõimet märkimisväärselt (sealsamas). Ühildumise piirangud on inimese kohustused viibida samal ajal mõnes kohas koos teiste inimestega, näiteks koolis või tööl käimine (sealsamas). Autoriteedi piirangud tulenevad ühiskondlikest reeglitest, mis piiravad isikute ligipääsu teatud kohtadele, näiteks ei saa inimesed alati piiramatult riigipiire ületada (sealsamas).

Hägerstrand (1970) tõi ajageograafiliste uuringute võimalike andmeallikatena välja liikumispäevikud, inimeste elujooned ning vaatlused. Mobiiliandmete eelsel ajal keskendusidki uuringud valdavalt inimgruppide uurimisele liikumispäevikute alusel (Schönfelder & Axhausen, 2003).

1.2 Mobiiliandmete kasutamine

20. sajandi teine pool tõi kaasa kiire tehnoloogilise arengu, mis kestab praeguseni. Areng oli ülimalt kiire nii inimeste liikumist võimaldavas transpordivaldkonnas kui ka informatsiooni liikumist toetavas infotehnoloogia valdkonnas. Juba 2006. aastal kuulutasid Sheller ja Urry (2006), et sotsiaalteadustes on kätte jõudnud uus paradigma, mis seab keskele kohale inimeste liikumise uurimise uute moodsate vahenditega. Samuti tõid nad esile uute andmeallikate esiletõusu ning arvutite mõju tähtsust (Sheller & Urry, 2006).

Infotehnoloogia areng võimaldas lisaks Hägerstrandi pakutud andmeallikatele ajageograafilistes uuringutes kasutusele võtta ka uut tüüpi andmeid – mobiiliandmeid (Ahas *et al.*, 2010). Tegemist on asukohaandmetega, mis tekivad mobiiltelefoni kasutamisel ning nende hulka kuuluvad nii passiivse mobiilpositsioneerimise (CDR, ingl. k *Call Detail Record*), aktiivse mobiilpositsioneerimise kui ka GPSi (*Global Positioning System*) või üldisemalt GNSSi (ingl. k *Global Navigation Satellite System*) põhiselt kogutud andmed.

Passiivse mobiilpositsioneerimise andmed koosnevad kõnetoimingute ja/või internetikasutuse logidest, mille toimumiskoht on määratud mobiilimasti täpsusega, kus ei ole võimalik seostada isikut ja andmeid (Roosaare *et al.*, 2019). Aktiivne mobiilpositsioneerimine on sarnane passiivsele, kuid andmed põhinevad mobiilioperaatori poolt tehtavatel päringutel (Ahas *et al.*, 2010), on seotud kindla isikuga ja on suurema täpsusega kui passiivne mobiilpositsioneerimine (Roosaare *et al.*, 2019). GPS-andmed on eelmisest kahest meetodist täpsemad, kuid samas ka energiakulukamad ning nõuavad eraldiseisva andmeid koguva rakenduse kasutamist.

Tänu mobiiltelefonide laialdasele levikule ühiskonnas on tegemist väga laiapõhjalise andmestikuga, mille kombineerimisel teist tüüpi andmetega on võimalik uurida väga erinevaid aspekte. Mobiilpositsioneerimise andmeid on kasutatud muuhulgas linnade uurimiseks (Ahas *et al.*, 2010, 2015), turismiuuringutes (nt Raun, 2020), hädaolukordadele lahenduste planeerimises (Tominga *et al.*, 2023), inimeste ööpäevarütmi uurimises (Kurvits, 2020; Linde, 2016), ühistranspordi uurimises (Haamer, 2022), keskkonnale eksponeerituse uurimises (Jankowska *et al.*, 2023) ning paljudes teistes teadusvaldkondades.

Lisaks teadustööle nähakse mobiiliandmetel suurt potentsiaali ka ametliku statistika loomisel. Eesti Pank võttis juba 2008. aastal mobiiliandmed kasutusele riikliku turismistatistika koostamisel eelistades neid turistide hulga hindamisel näiteks majutusasutuste ööbimisandmetele (Saluveer *et al.*, 2020). Ka Statistikaamet on seadnud eesmärgiks kaasata rohkem suurandmeid, sh mobiiliandmeid riikliku statistika koostamisse (Statistikaamet, 2018). Eesti eduloo eeskujul on sarnaseid võimalusi turismigeograafia koostamiseks otsinud ka Eurostat (Eurostat, 2014).

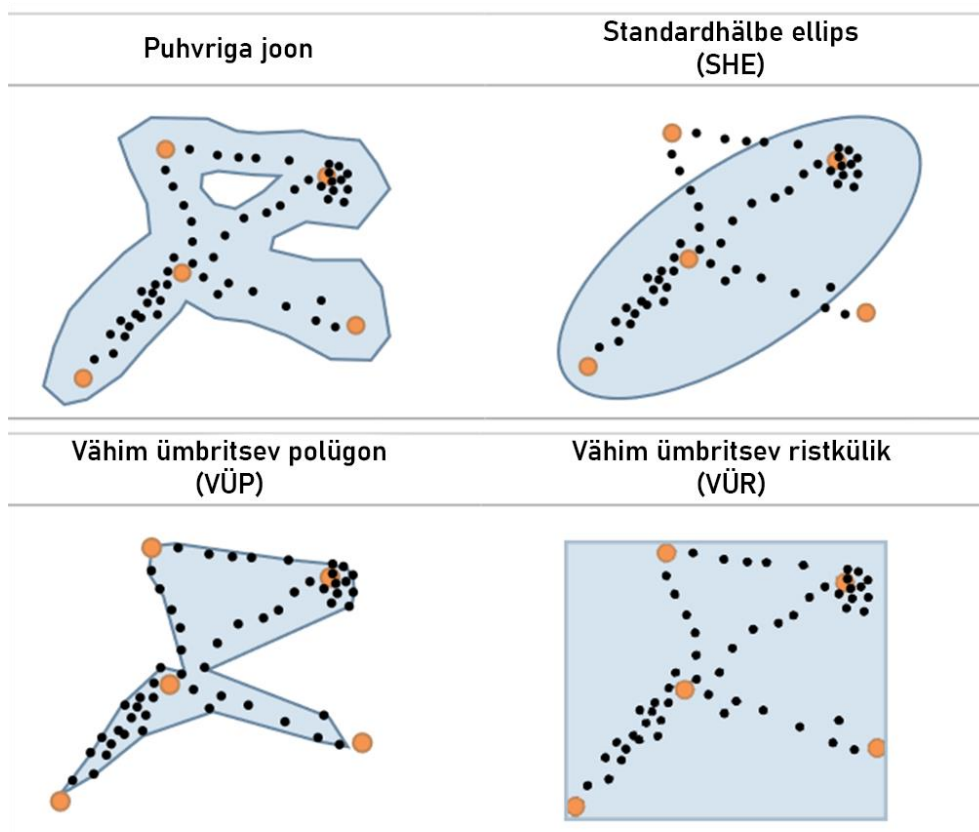
1.3 Ruumikujude kasutamine

Tihti ei eristata ingliskeelsetes tekstides tegevusruumi ja selle kirjeldamiseks kasutatavaid kujundeid ning kasutatakse mõlema kohta mõistet *activity space*. Mõnes tekstis on nende kujundite kirjeldamiseks kasutusel ka mõiste *shape* ning harva ka mõisted *activity space measure* või *activity space type*. Kuna autori arvates on selle töö kontekstis oluline selgelt

eristada tegevusruumi ning tegevusruumi kirjeldavaid kujundeid (ellips, polügon, ristkülik) siis selles töös on nende kujundite kohta kasutusel mõiste ruumikuju.

Tegevusruumide uurimisel on üheks keskseks probleemiks erinevate inimeste tegevusruumide võrdlemine (Schönfelder & Axhausen, 2003). Kuna iga inimene kasutab ruumi erinevalt ning ei järgi kindlaid reegleid ega mustrid on tegevusruumide võrdlemisel vaja andmeid üldistada ja lihtsustada.

Ruumikasutuse iseloomustamiseks on olemas mitmeid meetodeid, kuid üheks laialt levinumaks on ruumikujude kasutamine. Ruumikujude kasutamine loob lisaks arvilisele võrdlusalusele ka lihtsasti visualiseeritava tulemuse. Kasutatavaid ruumikujusid on mitmeid erinevaid. Smith *et al* (2019) poolt avaldatud ülevaateartiklis leiti, et valdavalt kasutatakse kolme kujundit - vähim ümbritsev polügon (VÜP, inglise keeles *minimum convex polygon*), standardhälbe ellips (inglise keeles *standard deviation ellipsis*) ja puhvriga liikumisteedekonda ehk puhvriga joont (inglise keeles *buffer*) ning punktide põhist interpoleerimist (Smith *et al*, 2019). Selles töös käsitletakse ruumikujuna veel ka vähimat ümbritsevat ristkülikut (VÜR, inglise keeles *bounding box*), mis tavaliselt on kasutusel peatuste tuvastamiseks GPS-andmetest, sellel eesmärgil on seda kasutanud näiteks Haamer (2022) ja Sturge *et al* (2021). Ruumikujusid on kujutatud ka joonisel 1, mis põhineb Smith *et al* (2019) joonisel, mida töö autor on muutnud asendades interpoleerimise meetodi VÜRiga.



Joonis 1. Ruumikujude näidised. Mustad punktid – GPS punktid, oranžid punktid – ankurpunktid, sinine kujund – tegevusruum. Originaaljoonis (Smith *et al*, 2019), autori kohandus.

Vähima ümbritseva polügoni (VÜP) näol on tegemist kõige väiksema võimaliku pindalaga hulknurgaga, milles sisalduvad kõik selle arvutamiseks kasutatavad punktid. Vähim ümbritsev ristkülik (VÜR) on oma olemuselt VÜPi erijuht ehk tegemist on nelinurgaga, mis sisaldab endas kõiki selle arvutamiseks kasutatud punkte. Puhvriga joon ühendab GPS-punktid kronoloogilises järjekorras joonega ning lisab selle ümber määratud suurusega puhvri. Standardhälbe ellipsi puhul jääb Smith *et al* (2019) ülevaateartikli kohaselt ellipsisse vastavalt arvutusmeetodile ühe või kahe standardhälbe ulatuses punkte. Sama ülevaateartikkel leidis, et puhvriga joone ruumikuju kasutamise puhul varieerus erinevates uuringutes kasutatud puhvri raadius 80 ja 500 meetri vahel (sealsamas).

1.3.1 Ruumikujude arvutamine

Ruumikujude näol on tegemist väga mitmekülgse meetodiga, mida saab rakendada mitmete erinevate lähteandmestike puhul.

Esimeseks võimalikuks lähteandmestikuks on liikumispäevikutega kogutud andmed. Liikumispäevikuid kasutasid Newsome *et al* (1998) kogutud tegevuskohtade põhjal tegevusruumi kirjeldavate ellipsite arvutamiseks. Teiseks võimalikuks andmeallikaks on passiivse mobiilpositsioneerimise andmed, mida kasutasid päevase ja ühe kuu tegevusruumi arvutamiseks Järv *et al* (2014), kes samuti leidsid esmalt inimeste jaoks olulised tegevuskohad ning seejärel arvutasid päevase ja ühe kuu tegevusruumi iseloomustavad standardhälbe ellipsid. Kolmanda andmeallikana on kasutatud GPSi abil kogutud andmeid, mille põhjal arvutatakse esmalt tegevuskohad ja nende põhjal ruumikujud (Gavela, 2020) või kasutatakse kõiki GPSiga kogutud punkte ruumikujude arvutamisel (Hirsch *et al*, 2014).

1.4 Varasemad üliõpilastööd

Tartu Ülikooli geograafia osakonnas on aastatel 2013-2023 kaitstud 15 bakalaureuse- ja magistritööd, mis on kasutanud lähteandmetena rakendusega MobilityLog kogutud GPS-andmeid. Kalamaja ja Priisle elanikke on nendest töödest uurinud neli magistritööd.

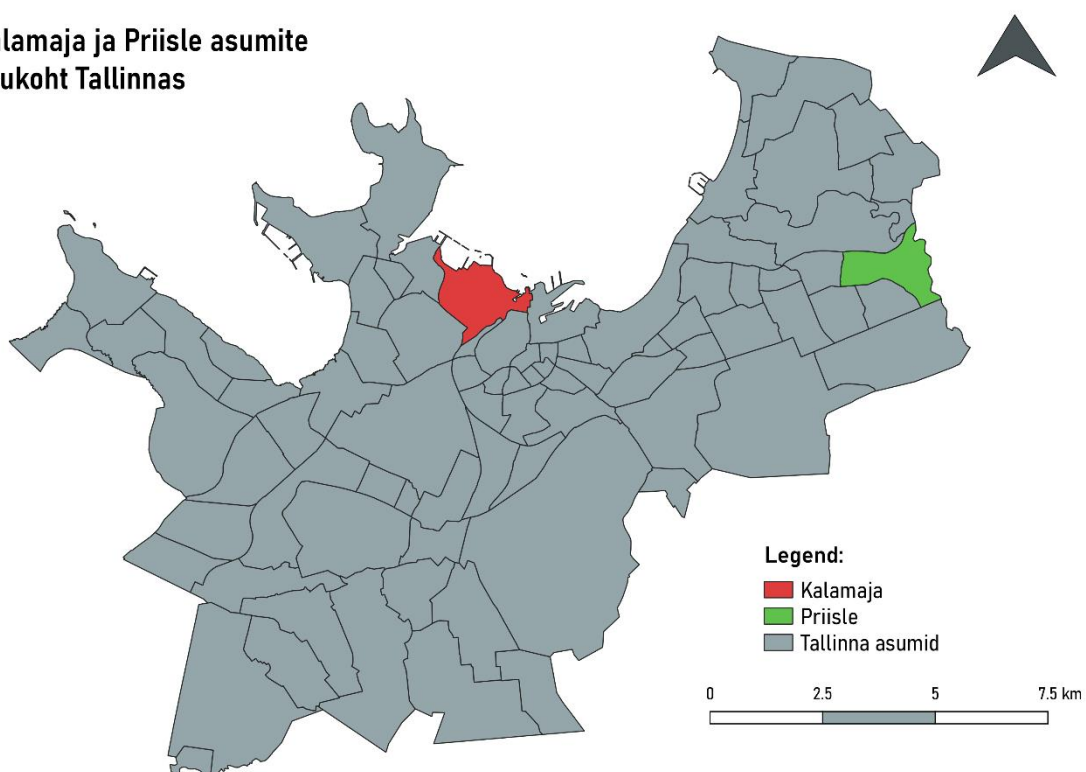
Zernant (2023) keskendus Kalamaja ja Priisle elanike looduslike alade külastamise ja sotsiaalsete tunnuste vaheliste seoste uurimisele. Paris (2022) uuris enda magistritöö raames etnilisuse rolli Kalamaja ja Priisle elanike kohtade külastamises. Haamer (2022) hindas Priisle ja Kalamaja piirkondade elanike ligipääsetavust ühistranspordile ning Kozlovskiy (2023) uuris elukohta, sotsiaalsete ja demograafiliste tegurite mõju inimeste tegevusruumidele.

2. Andmed ja töö käik

2.1 Andmete kogumine

Töö tegemise aluseks on andmed, mis koguti teadusprojekti “Ruumilise eraldatuse nõiarongi geograafiline analüüs” käigus, mille vastutav täitja on prof. Tiit Tammaru (projekti number PRG306). Projekti eesmärgiks oli uurida Tallinna näitel ruumilise eraldatuse tekkimist registritutitefonide- ja küsitlusandmete põhjal (Etis, i.a.). Projekt toimus ajavahemikus 1.01.2019–31.12.2023 ja seda rahastas Sihtasutus Eesti Teadusagentuur (sealsamas). Uuringus uuriti Tallinnas Kalamaja ning Priisle asumites elavaid inimesi. Uurimispiirkondade asukohti on kujutatud joonisel 2.

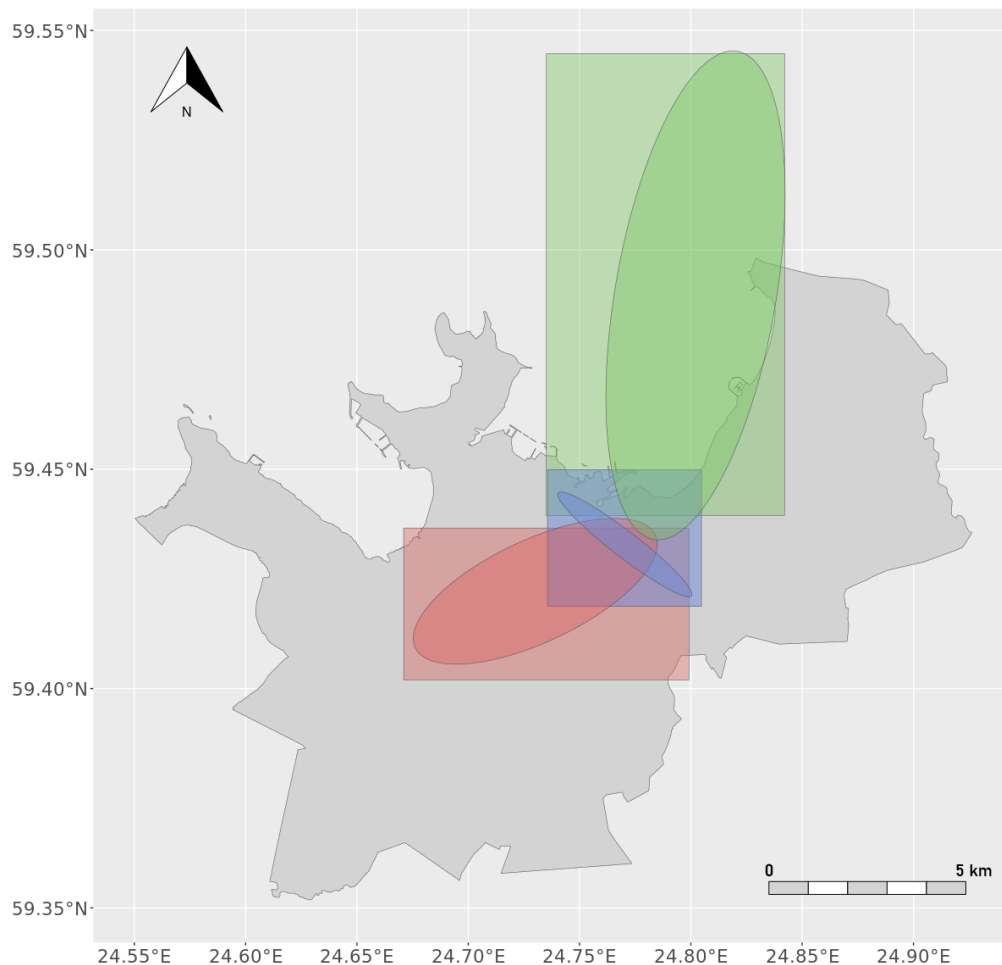
Kalamaja ja Priisle asumite asukoht Tallinnas



Joonis 2. Kalamaja ja Priisle asumite paiknemine Tallinna linnas. Algandmeteks on Tallinna asumiaandmed (02.06.2020) (Tallinna Linnaplaneerimise Amet, 2020).

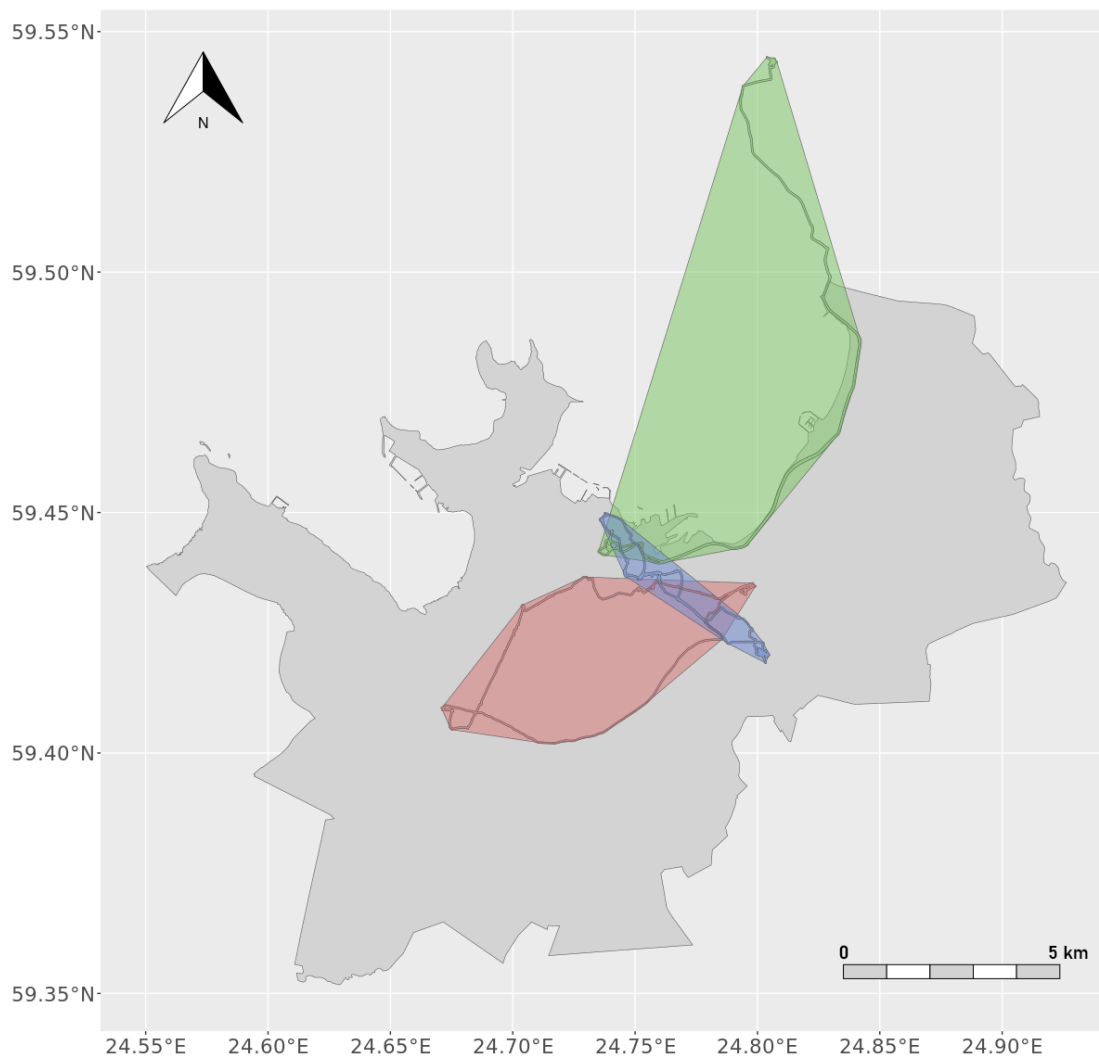
See töö põhineb teadusprojekti käigus kogutud GPS-andmetel, milleks kasutati Tartu Ülikooli mobiilsusuuringute labori poolt väljatöötatud MobilityLog rakendust. Projekti andmekogumine toimus 2020. aasta maikuust kuni 2022. aasta märtsikuuni. Samuti kasutatakse ka teadusprojekti käigus uuringus osalejate kohta kogutud sotsiaal-demograafilisi andmeid.

Käesoleva uurimistöö raames ei kasutata mitte projekti raames kogutud toorandmeid (GPSi poolt salvestatud asukohapunktid), vaid GPS-andmetest arvutatud ruumikujusid. Ruumikujude arvutamiseks kasutati vaid neid GPS-punkte, mis jäid Eesti Vabariigi territooriumile. Ruumikujud arvutas kaasprofessor Anto Aasa, kes on ka käesoleva töö juhendaja. Töös analüüsitakse nelja erinevat ruumikuju. Joonisel 3 on kujutatud kolme uuringus osalenu ühe päeva jooksul tehtud liikumiste põhjal arvutatud vähimat ümbritsevat ristkülikut (VÜR) ning standardhälbe ellipsit (SHE).



Joonis 3. Kolme uuringus osalenud inimese ühe päeva tegevusruum kujutatuna standardhälbe ellipsi ja vähima ümbritseva ristkülikuga. Aluskaardiks on haldus- ja asustusjaotus 01.05.2024 seisuga (Maa-amet, 2024).

Joonisel 4 on kujutatud samade inimeste sama päeva liikumiste põhjal arvutatud vähimat ümbritsevat polügoni (VÜP) ning puhvriga joont. Puhvri raadiuseks on selles töös 20m, mis valiti seetõttu, et sellisel raadiuse puhul kujutab puhvriga joon inimese poolt vahetult kogetavat elukeskkonda.



Joonis 4. Kolme uuringus osalenud inimese ühe päeva tegevusruum kujutatuna puhvriga joone ning ja vähima ümbritseva polügoniga. Aluskaardiks on haldus- ja asustusjaotus 01.05.2024 seisuga (Maa-amet, 2024).

2.1.1 Andmete kirjeldus

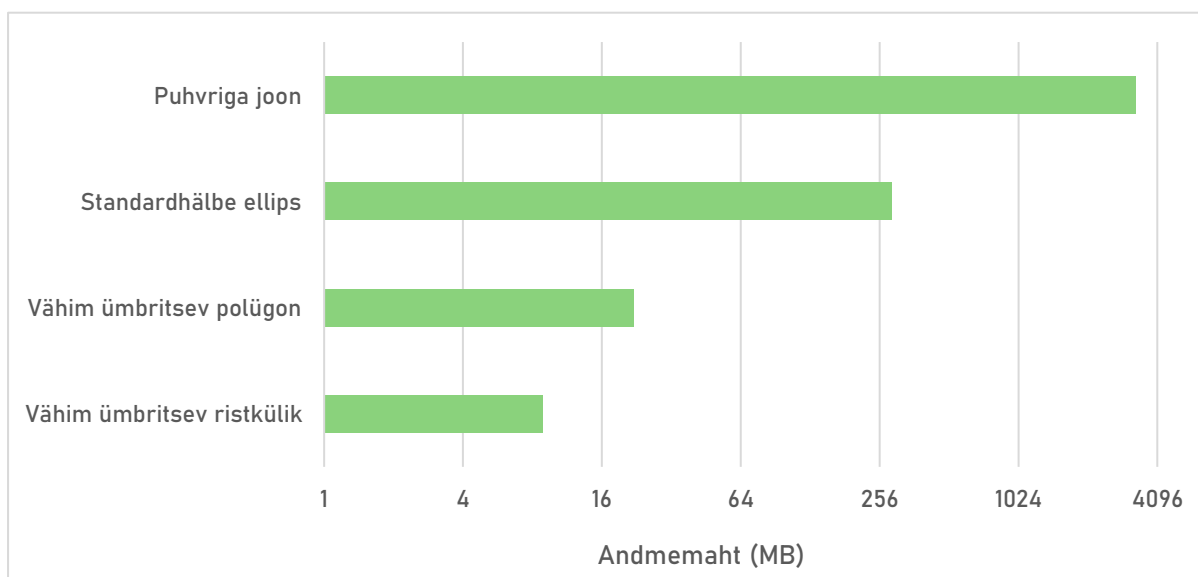
Teadusprojekti “Ruumilise eraldatuse nõiarangi geograafiline analüüs” käigus kogutud andmetest kasutakse selles töös andmeid, mis laekusid ajavahemiku 19. oktoober 2020 kuni 18. oktoober 2021. Sellesse ajaperioodi jääb iga analüüsitava ruumikuju kohta 25 226 kuni 25 996 andmerida. Erinevus ruumikujude andmeridade arvus tuleneb sellest, et ruumikujude arvutamiseks on vaja erinevat arvu asukohapunkte. VÜPi ja VÜRi arvutamiseks on vaja vähemalt ühte andmepunkti, puhvriga joone arvutamiseks on vaja vähemalt kahte andmepunkti ning standardhälbe ellipsi arvutamiseks vähemalt kolme asukohapunkti. Seetõttu on väiksema miinimumpunktide arvuga meetoditel rohkem andmeridasid, kui suurema miinimumpunktide arvuga meetoditel. Ruumikujude jaotumine on ära toodud tabelis 1.

Tabel 1. Töös analüüsitavate andmeridade jaotumine erinevate ruumikujude vahel.

Ruumikuju	Andmeridasid enne filtreerimist	Töö analüüsis kasutatud andmeridade arv
Vähim ümbritsev ristkülik (VÜR)	37 954	25 996
Vähim ümbritsev polügon (VÜP)	37 954	25 996
Puhvriga joon	37 085	25 338
Standardhälbe ellips	37 006	25 226

On oluline välja tuua, et lisaks kuupäevapõhisele piiramisele eemaldati andmestikust ka andmereal, mille edastanud kasutajatunnuse keelevalikuks oli märgitud „Other“. Põhjuseks oli see, et selliseid inimesi oli vähe ehk kaheksa ning nende kaasamine andmeanalüüsi ei oleks olnud seetõttu sobilik. Samuti ei ole võimalik olla kindel, et tegemist ei ole andmeveega. „Other“ keelega kasutajate andmeridade eemaldamine moodustas kõikidest eemaldatud andmeridades sõltuvalt ruumikujust 1898 kuni 1963 andmerida.

Tulenevalt sellest, et erinevad ruumikujud on geomeetriliselt erineva keerukusega on ka nende hoiustamiseks vajalik andmemaht erinev. Joonisel 5 on toodud välja ühe ruumikuju hoiustamiseks vajalik andmemaht erinevate ruumikujude korral. Joonise puhul tasub tähele panna, et andmemahtude telg on logaritmiline. Joonis näitab selgelt, et puhvriga joone on teistest tegevusruumi kujutusviisidest kordades suurema andmemahuga, samas kui vähim ümbritsev polügon ja vähim ümbritsev ristkülik on peaaegu sama väikse andmemahuga.



Joonis 5. Ühe ruumikuju hoiustamiseks vajalik andmemaht

Töös kasutatavad ruumikujud on arvatud 184 erineva isiku andmete põhjal. Iga andmerida sisaldas infot, mis kuupäeval ja millisel kasutajatunnusel andmed laekusid, samuti ruumikuju iseloomustavate näitajate väärtusi ning kasutajatunnusega seotud sotsiaalsete tunnuste väärtusi. Lähteandmetes olid kajastatud osaleja elukoht (kas Kalamaja või Priisle), vanus, haridustase, sissetulek, ameti tüüp ning leibkonna suurus.

Selle töö käigus analüüsiti ruumikujude seost kolme sotsiaal-demograafilise tunnusega, milleks valiti emakeel, elukoht ja haridustase. Sellise valiku põhjuseks on nende tunnuste olulisus ruumikasutusega seotud uuringutes, eriti uuringutes, mis keskenduvad segregatsioonile.

Emakeele alusel moodustasid 54,9% valimist eesti keelt kõnelejad ning 45,1% vene keelt kõnelejad (vastavalt 101 ja 83 inimest). Elukoha alusel oli valimis 52,2% Priisle elanikke (96 inimest) ning 47,8% Kalamaja elanikke (88 inimest). Haridustaseme puhul ei olnud valim nii tasakaalus, kõrgemalt haritud inimesed moodustasid 62,5% (115 inimest) ning madalamalt haritud 37,5% (69 inimest).

2.1.2 Rakendus MobilityLog

MobilityLog on Androidi operatsioonisüsteemiga nutitelefonidel töötav mobiiltelefonirakendus (IMO, 2023). Rakenduse eesmärgiks on koguda uuringutes osalejate kohta mitmesuguseid andmeid. Rakenduse kaudu kogutakse osalejatelt kolme tüüpi andmeid – asukohaga seotud koordinaadid ja kellaajad, kõnetoimingute andmed ja telefonikasutusega seotud andmed (sealsamas). Lisaks on uuringu läbiviijatel võimalik esitada uuringus osalejale asukohaspetsiifilisi küsimusi.

MobilityLog salvestab inimeste asukohta erineva tihedusega, et vähendada väga suurt energiakulu ning seeläbi ka telefoni aku tühjenemist (Poom, 2019). MobilityLog algatab seadme asukoha salvestamise järgnevate olukordade puhul: kui telefon tuvastab 10 sekundiliste intervallidega 3 korda järjest telefoni liikumise või kui viimasest asukoha tuvastamisest on möödas rohkem kui 15 minutit (sealsamas). Liikumise ajal salvestatakse asukohti sõltuvalt liikumiskiirusest kas iga sekundi tagant või kui liikumine on aeglasem kui kolm m/s, siis iga 16 sekundi tagant (sealsamas).

2.1.3 Andmekaitse

Töös kasutatud andmed, tehti töö läbiviijale saadavaks Tartu Ülikooli mobiilsusuuringute labori poolt andmekasutuslepingu alusel. Andmekasutusleping 6-3/OM/775 on leitav Tartu Ülikooli dokumendihalduse süsteemis. Kõik andmed tehti töö autorile kättesaadavaks pseudonüümituna ehk isikuandmed olid asendatud genereeritud ID numbriga, kus ühele

inimesele vastab üks genereeritud number. Andmete väljastamine kooskõlastati ka Tartu Ülikooli eetikakomiteega, kes lisas selle töö originaalprojekti eetikaloale (eetikaloa number 298T-12).

Andmete töötlemine toimus Tartu Ülikooli mobiilsuuringute labori serveris ilma võrguühendusega virtuaalarvutis, et vältida tundlike andmete lekkimist. Andmetele tagati ligipääs kasutajakonto ja parooli alusel Tartu Ülikooli võrgust või virtuaalse turvalise ühenduse (inglise keeles *virtual private network*) vahendusel.

2.2 Andmetöötlus ja -analüüs

Andmetöötluse esimeseks sammuks oli andmetest valitud ajavahemikku kuuluvate andmete väljafiltreerimine kuupäeva alusel. Ühe aasta pikkune ajavahemik valiti, sest selles ajavahemikus laekus igapäevaselt andmeid rohkem kui 50 inimeselt (minimaalselt 54), selles ajavahemikus valdav osa uuringu andmetest ning ühe aasta pikkust ajaperioodi on varasemalt kasutatud tegevusruumide uurimisel (näiteks Järv *et al*, 2014) ning GPSi põhistel uuringutel (näiteks Millward *et al*, 2019).

Seejärel arvutati korrelatsioonimaatriks ruumikujude erinevatele näitajate vahel, et mõista üldisemalt andmestikku ning ruumikujude omavahelisi seoseid. Lisaks koostati ülevaade töö valimist, andmete laekumisest ning ruumikujudest.

Andmetest ülevaate koostamisel leiti, et andmetes on suurel määral erandeid ehk esines hulgaliselt ruumikujusid, mille pindalade väärtused olid ebatavaliselt suured. Sellest tulenevalt tekkis kahtlusi andmete normaaljaotusele vastavuse osas, mistõttu kontrolliti iga ruumikuju pindala ja übermõõdu väärtuste vastavust normaaljaotusele kasutades selleks Lilliefors'i testi ning Pearsoni hii-ruut testi. Kõikide ruumikujude puhul olid mõlema tunnuse testide $p < 0,01$ ehk andmed ei vasta normaaljaotusele. Seetõttu kasutati andmete keskmise väärtuse iseloomustamiseks mitte keskmist vaid mediaanväärtust ning jaotuste vaheliste erinevuste tuvastamiseks mitteparameetrilisi statistilisi teste.

Järgmise sammuna koostati väiksemad valimid erinevate ruumikujude kohta, grupeerides andmeid vastavalt valitud tunnustele (emakeel, elukoht ja haridustase). Näiteks oli üheks selliseks valimiks kõik eesti keelt kõnelevate inimestega seotud ellipsid. Peale filtreerimist leiti iga grupi kohta mediaanväärtused, mida võrreldi teiste gruppide mediaanväärtustega.

Andmetöötluse osana kasutati Duncani testi, et leida, kas tunnuste ja ruumikujude põhjal tekkinud grupid on üksteisest sama tunnuse sees statistiliselt erinevad (nt Priisle vs Kalamaja elanike VÜPid).

Analüüsi viimase sammuna kontrolliti, kas grupeerimiseks kasutatud tunnused on statistiliselt oluliselt. Selleks kasutati Kruskal-Wallise testi, mis valiti seetõttu, et tegemist on mittestandardjaotusega valimite jaoks sobiliku testiga. Kruskal-Wallise testi tulemusena selgus, et analüüsitud tunnused (emakeel, elukoht ja haridustase) on statistiliselt olulised, sest kõikide valimite testide tulemuseks oli olulisuse tõenäosuse (p) väärtus alla 0,001.

Andmete analüüsimiseks kasutati arvutiprogrammi RStudio Server 2023.06.0 Build 421 koos pakettidega agricolae, dbplyr, ggspatial, lubridate, nortest, readr, RPostgreSQL, sf, tidyverse, ja tmap. RStudio Server'i, QGIS 3.16 ja Adobe Photoshop 2024 abil valmisid joonised ning kaardid. Lihtsamate arvutuste tegemiseks ning tabelite ja jooniste koostamiseks kasutati ka Microsoft Excelit.

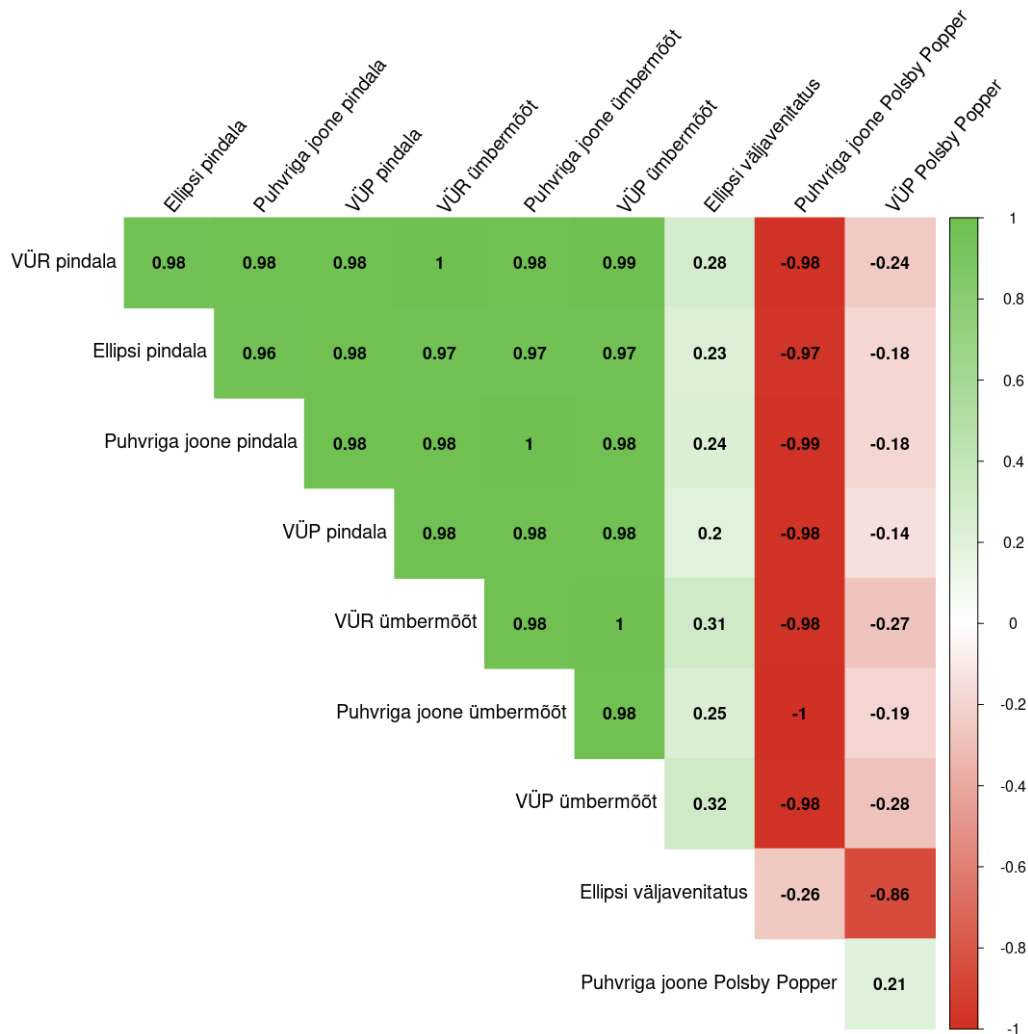
2.3 Tehisintellekti kasutamine

Töö koostamisel kasutati tehisintellekti minimaalselt. Tehisintellekti kasutati võõrkeelsete mõistete tõlkimiseks ning lahtimõtestamiseks, koodi kirjutamisel tekkinud vigade lahendamiseks ning koodikirjutamisel näidete genereerimiseks. Oluline on välja tuua, et koodi kontrollimisel ei sisaldanud tehisintellektile sisestatud koodiread andmeid, vaid ainult muutujate tähiseid või muud mittetundlikku infot. Töö käigus kasutati tehisaru ChatGPT3.5 ja Microsoft Copilot.

3. Tulemused

3.1 Korrelatsioonimaatriks

Spearmani korrelatsioonimaatriksis (joonis 6) võrreldi töös analüüsitud nelja ruumikuju (VÜP, VÜR, standardhälbe ellips ning puhvriga joon) pindalade, ümberrõõtude ning väljavenitatusesega seotud näitajate vahelisi korrelatsioone.



Joonis 6. Spearmani korrelatsioonimaatriks erinevate ruumikujusid iseloomustavate tunnuste vahel

VÜRi, VÜPi, ellipsi ja puhvriga joone pindalade vahel on väga tugev seos (korrelatsioonikordajad vahemikus 0,96 kuni 0,98), mis kinnitab Sila-Nowicka *et al* (2023) uuringus saadud tulemust, mis näitas, et standardhälbe ellipsi ja VÜPi pindalade vaheline korrelatsioon oli 0,93. Ootuspäraselt on omavahel samuti tugevas seoses ka VÜPi, VÜRi ja puhvriga joone ümbermõõdud, kus korrelatsioonikordajad on vahemikus 0,93 kuni 0,99.

Sarnaselt pindalaga on väga tugev korrelatsioon ka VÜRi, VÜPi ja puhvriga joone ruumikujude ümbermõõtude vahel, kus korrelatsioonikordajad jäävad vahemikku 0,98 kuni 1.

Kolmas grupp analüüsitud tunnuseid kirjeldavad ruumikujude väljavenitatust. Polsby-Popperi kompaktsus (inglise keeles *Polsby-Popper compactness*) on tunnus, mis kirjeldab kujundi pindala ja ümbermõõdu suhet läbi valemi

$$PP = \frac{4\pi S}{P^2}, \text{ kus } S - \text{kujundi pindala, } P - \text{kujundi ümbermõõt (Polsby \& Popper, 1991).}$$

Selle näitaja väärtused jäävad vahemikku null kuni üks ning mida suurem on väärtus, seda enam sarnaneb kujund ringile ehk seda kompaktsem ta on (sealsamas). Selle omaduse tõttu on ka Polsby-Popperi kompaktsusega seotud korrelatsioonid valdavalt negatiivsed ehk näiteks, mida suurem on puhvriga joone pindala, seda vähem sarnaneb kujund ringile ning seda väiksem on Polsby-Popperi kompaktsuse näitaja.

Teiseks ruumikuju väljavenitatust iseloomustavaks näitajaks on *eccentricity* ehk ekstsentrilisus. Selles töös on kasutatud seda standardhälve ellipsi iseloomustamiseks. Ekstsentrilisust arvutatakse ellipsi puhul valemiga

$$\sqrt{1 - \frac{b^2}{a^2}},$$
 a - ellipsi pikema telje pikkus, b - ellipsi lühema telje pikkus.

Ekstsentrilisus jääb ellipsite puhul vahemikku 0 kuni 1, kus väärtus 0 tähistab perfektset ringi ehk ekstsentrilisuse puudumist.

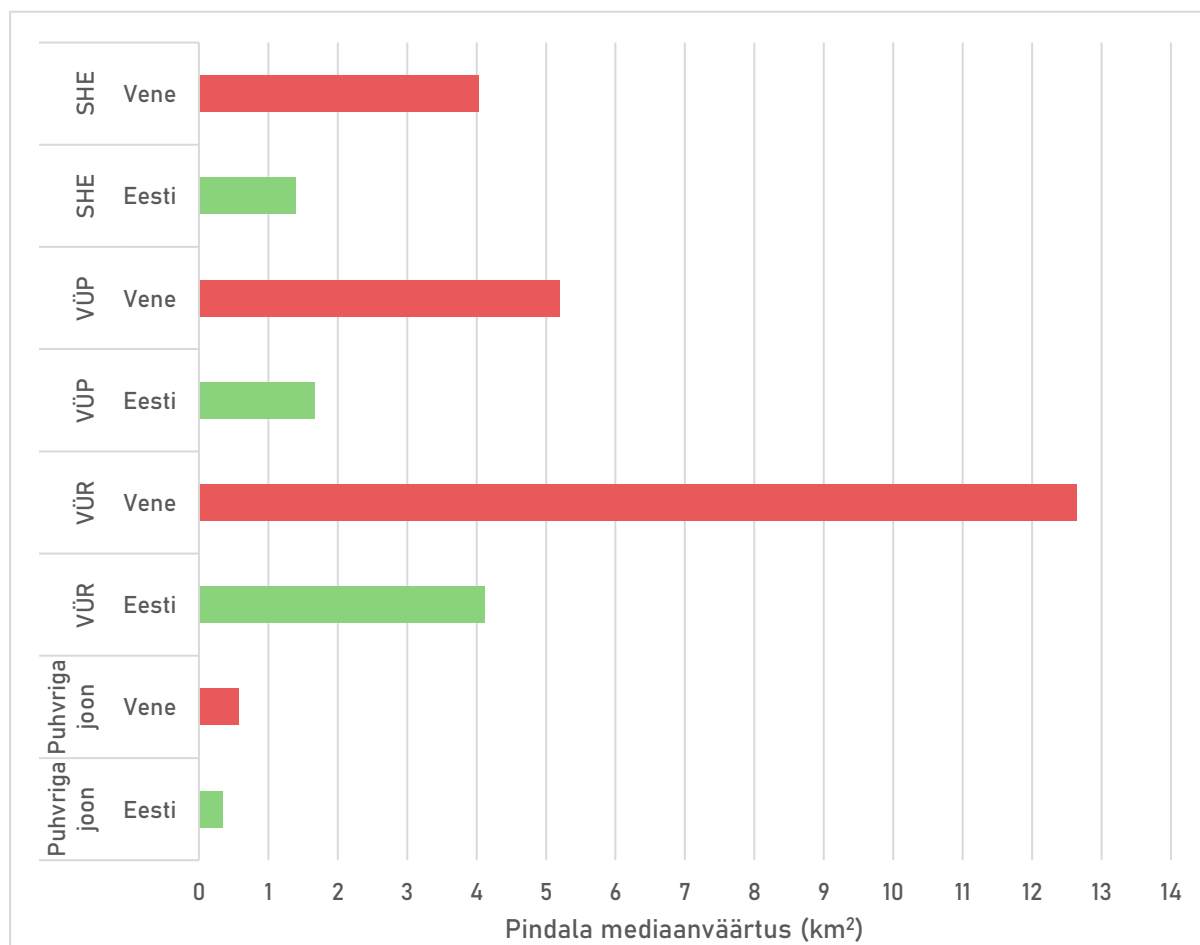
Väljavenitatust käsitlevatest näitajatest ilmnes ainus tugev korrelatsioon standardhälve ellipsi ekstsentrilisuse ja vähima ümbritseva polügooni (VÜP) Polsby-Popperi kompaktsuse vahel, mis on ootuspäraselt negatiivne korrelatsioon (kordaja -0,86).

Ruumikujude siseselt on Polsby-Popperi kompaktsuse ja pindala ning Polsby-Popperi kompaktsuse ja ruumikuju ümbermõõtude vahel negatiivsed korrelatsioonid – puhvriga joonel vastavalt -0,99 ning -1 ehk väga tugev korrelatsioon ja vähimal ümbritseval polügonil vastavalt -0,14 ning -0,28 ehk nõrk korrelatsioon. Standardhälve ellipsi pindala ja ekstsentrilisuse vahel aga on nõrk positiivne korrelatsioon (kordaja 0,23).

3.2 Mediaanväärtuste võrdlus

Ruumikujude omavaheliseks võrdlemiseks arvutati andmeanalüüsi käigus kõikide ruumikujude pindalade ning VÜPi, VÜRi ja puhvriga joone ümbermõõtudele mediaanväärtused. Mediaanväärtuste arvutamisel esmalt grupeeriti ruumikujud vastavalt kas emakeele, elukoha või hariduse alusel, seejärel arvutati kahe grupi (nt Kalamaja ja Priisle elanike) vaheline mediaanväärtuste erinevus. Erinevate ruumikujude pindalade võrdlus on toodud lisas 1 tabelis 4 ning ümbermõõtude võrdlus on toodud lisas 1 tabelis 5. Tulenevalt sellest, et kõikidele ruumikujudel polnud lähteandmetes ümbermõõtu arvutatud loobuti ümbermõõdu kasutamisest võrdlustes ning järgnev analüüs keskendus vaid pindalade analüüsimisele.

Kõik neli ruumikuju näitavad, et emakeele alusel võrreldes on ruumikujude pindalade mediaanid venekeelse emakeelega inimestel suuremad, kui eestikeelse emakeelega inimestel (joonis 7).

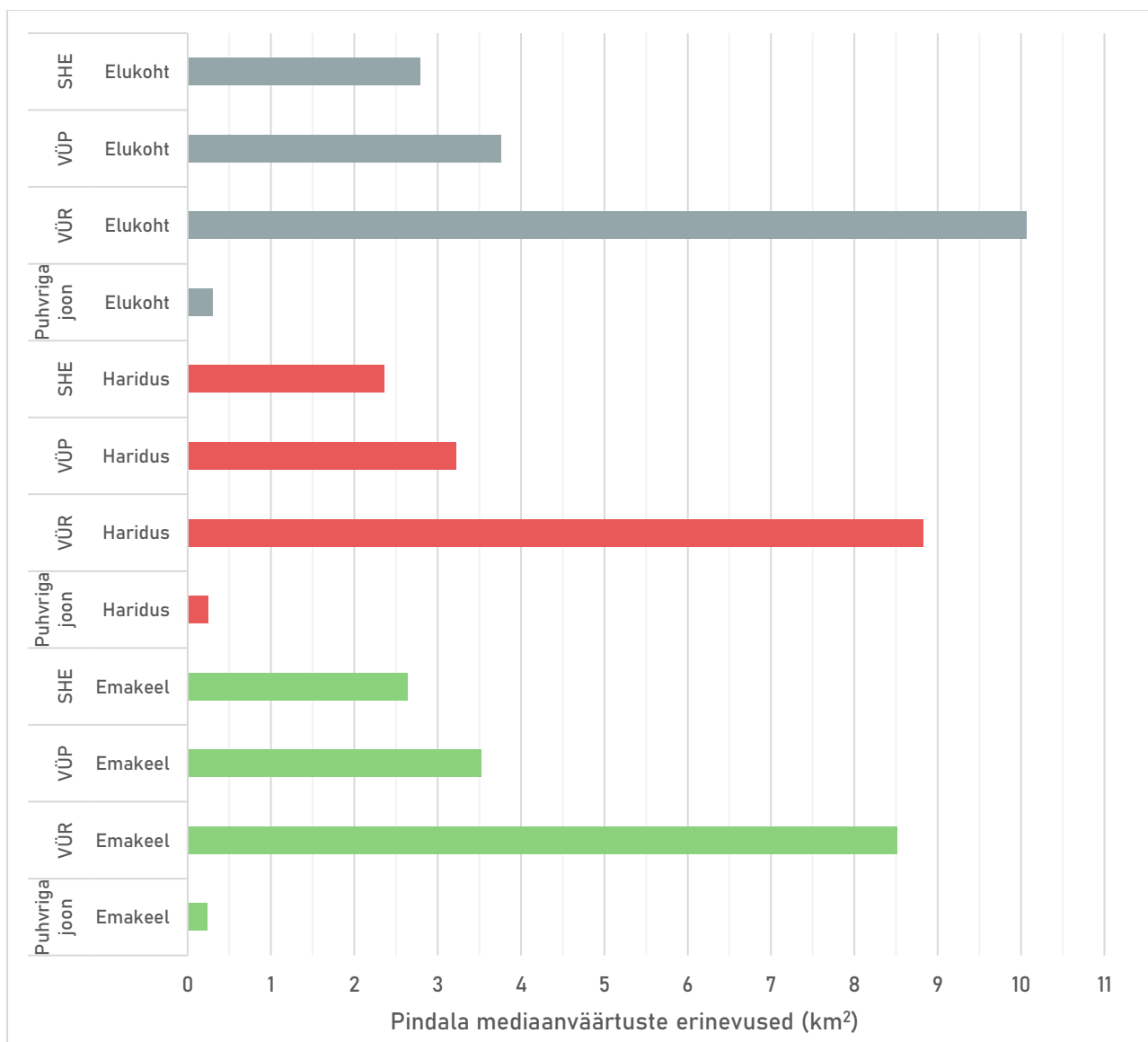


Joonis 7. Erinevate ruumikujude pindalade mediaanväärtused grupeerituna isiku emakeele põhjal

Haridustasemete alusel võrreldes näitavad kõik ruumikujud, et madalama haridustasemega inimestel on ruumikujude pindalade mediaanväärtus suurem, kui kõrgema haridustasemega inimestel. Elukohtade võrdlusel näitasid kõik meetodid, et Priisle asumi elanike tegevusruumide pindalade mediaanväärtus on suurem, kui Kalamaja asumi elanikel.

Ruumikujude mediaanide võrdlusest selgub, et puhvriga joone pindala on teiste meetoditega võrreldes 10 korda väiksema pindalaga, kui teised analüüsitud meetodid. Pindala väärtus on muudetav puhvri raadiuse muutmisega, selles töös on kasutusel 20 meetrise raadiusega puhver.

Tulenevalt teistest meetoditest märkimisväärselt väiksematest tegevusruumi pindalade väärtustest on puhvriga joonel kõikide võrreldavate tunnuste lõikes kõige väiksemad pindalade mediaanide erinevused (joonis 8).



Joonis 8. Ruumikujude pindalade mediaanväärtuste erinevused grupeerituna ruumikujude ning võrdlustunnuste alusel.

Kõigi kolme uuritud tunnuse alusel oli suurima pindalade mediaanide erinevusega ruumikujuks vähim ümbritsev ristkülik, mis suurim erinevus on elukohapõhisel võrdlemisel, kus mediaanide erinevus ulatus lausa 10,058 km². Teist kõige suuremat erinevust mediaanide vahel näitab vähim ümbritsev polügon, mille puhul jäävad erinevused vahemikku 3,215 km² kuni 3,753 km². Standardhälbe ellips (SHE) mediaanide erinevused jäävad samasse suurusjärku kui VÜPi omad olles vahemikus 2,360 km² kuni 2,789 km².

Eelmise kolme meetodiga võrreldes on märkimisväärselt väiksema pindalade mediaanide erinevusega puhvriga joon, mis on teiste ruumikujude pindalade erinevustest väiksem nii emakeele, haridustaseme kui ka elukohapõhisel võrdlemisel. Emakeele puhul on erinevus vaid

0,23 km², haridustasemete võrdluses oli erinevus 0,245 km² ja elukoha alusel võrreldes oli erinevus 0,230 km².

Tulenevalt sellest, et puhvriga joone ruumikuju pindala on märkimisväärselt väiksem, kui teised uuritud ruumikujud, siis leiab töö autor, et oleks kohatu piirduda ainult absoluutväärtuste võrdlemisega. Seda mõtet toetab ka fakt, et puhvriga joone puhul saab puhvri raadiust vabalt muuta, mis koheselt muudaks ka saadava ruumikuju pindala märkimisväärsel määral. Selleks, et viia erineva suurusega ruumikujud ühele võrreldavale alusele arvutati igale ruumikuju ja võrreldava tunnuse paarile niinimetatud võrdluskordaja. Võrdluskordaja on väärtus, mis on leitav valemiga

$k = \frac{a}{b}$, a - võrdluspaari suurem mediaanpindala, b - võrdluspaari väiksem mediaanpindala.

Võrdluskordajate väärtused on lisatud lisas 1 tabelitesse 4 ja 5 ning tuhandiku täpsusega on need toodud ka tabelis 2.

Tabel 2. Erinevate ruumikujude pindalade mediaanväärtuste võrdluskordajad ning summaarne võrdluskordaja

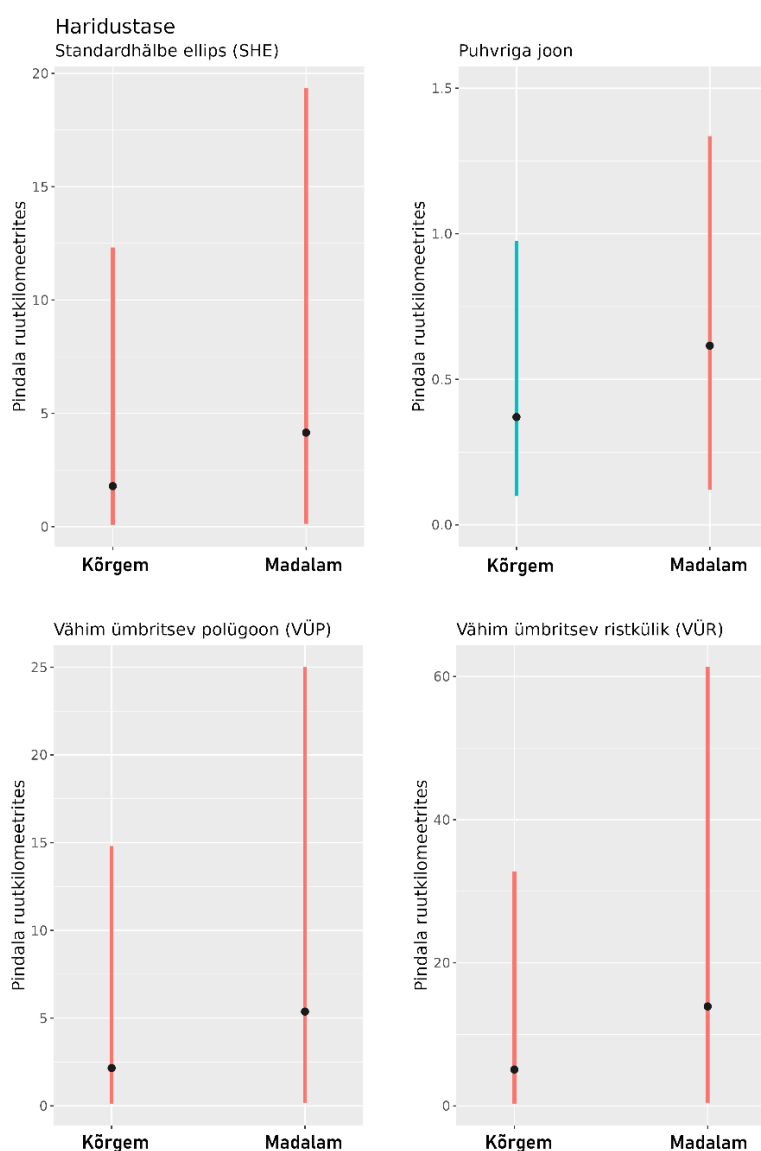
Tunnus	Puhvriga joon	VÜR	VÜP	Standardhälbe ellips
Emakeel	1,683	3,065	3,107	2,890
Haridus	1,661	2,734	2,493	2,320
Elukoht	2,016	4,189	3,737	3,423
Summa	5,360	9,988	9,337	8,633

Võrdluskordajate analüüs näitab selgelt sama trendi, mis absoluutväärtuste võrdlemine. Jätkuvalt on puhvriga joonel kõige väiksem erinevus pindalade mediaanide vahel. See trend kehtib kõigi kolme uuritud tunnuse puhul. Ülejäänud kolm meetodit on võrdlemisi sarnaste võrdluskordajate väärtustega. Võrreldes omavahel VÜRi, VÜPi ja standardhälbe ellipsi võrdluskordajaid on VÜRi kordajad teistest veidi suuremad, VÜPi omad keskmised ning ellipsi omad natukene väiksemad. Tabelis 2 on välja toodud ka summaarne võrdluskordaja, mis on kolme eri tunnuse kohta arvatud võrdluskordajate summa ning see näitab üldist trendi erinevate ruumikujude võrdluskordajate vahel.

3.3 Tunnusepõhine võrdlemine

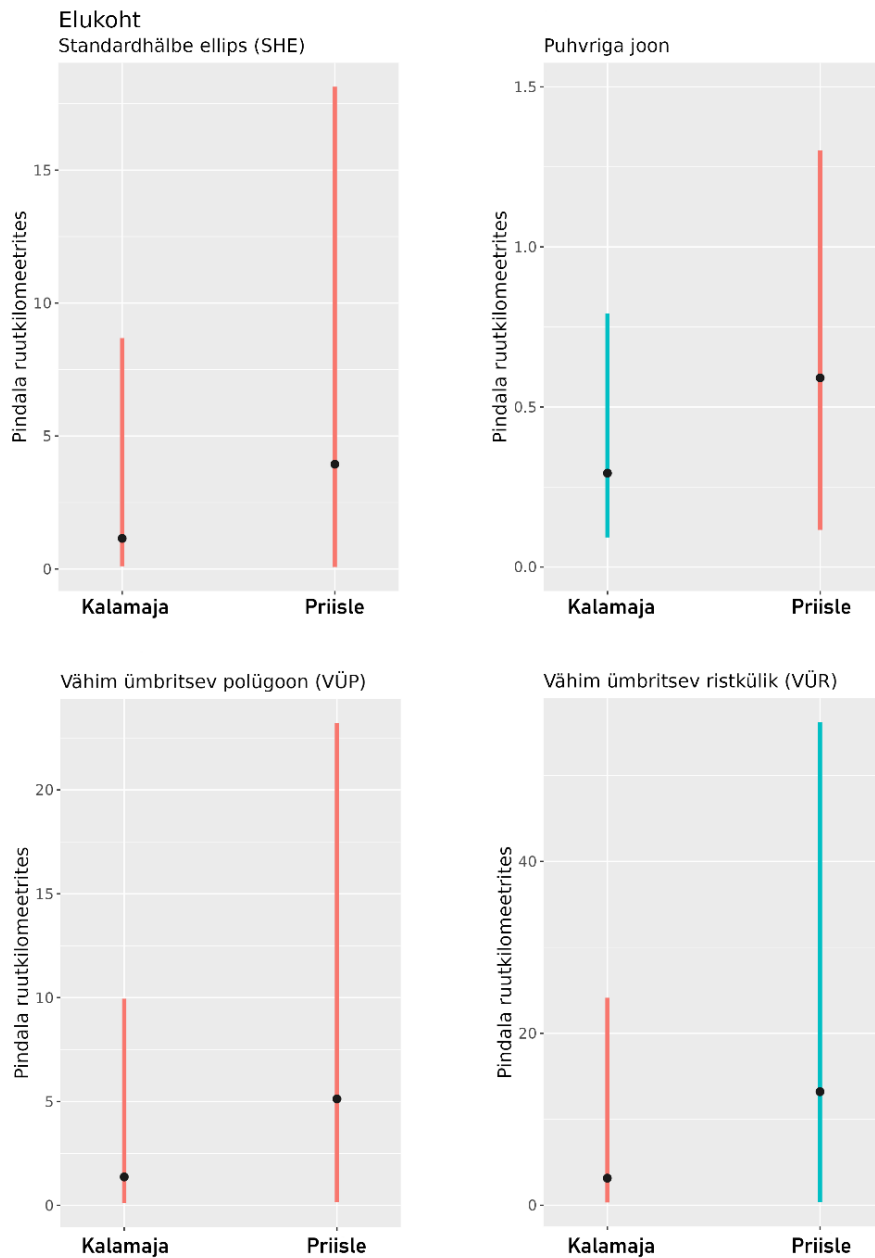
Mediaanide absoluutväärtuste võrdlemise visualiseerimiseks ja tunnusesiseste erinevuste statistiliseks kontrollimiseks kasutati esmalt Duncani testi.

Duncani test võrdleb valimeid mediaanide alusel ning jagab võrreldud valimid gruppidesse, kus sama grupi valimite vahel on seos ehk need valimid ei ole üksteisest statistiliselt erinevad. Selles töös kasutati valimitena ühe ruumikuju ning ühe sotsiaalse tunnuse kindlat väärtust, näiteks kõik puhvriga jooned, mis pärinevad kõrgharidusega inimestelt. Duncani test leidis, et haridustasemepõhiselt esineb statistiliselt oluline erinevus ainult puhvriga joone ruumikuju puhul, teiste ruumikujude juures statistiliselt olulist erinevust ei tuvastatud (joonis 9).



Joonis 9. Duncani testi tulemused haridustaseme põhjal, kujutatud modifitseeritud karpdiagrammidena. Joon tähistab kvartiilihaaret, punkt mediaani. Sama värvi jooned määrati Duncani testi poolt ühte gruppi kuuluvaks.

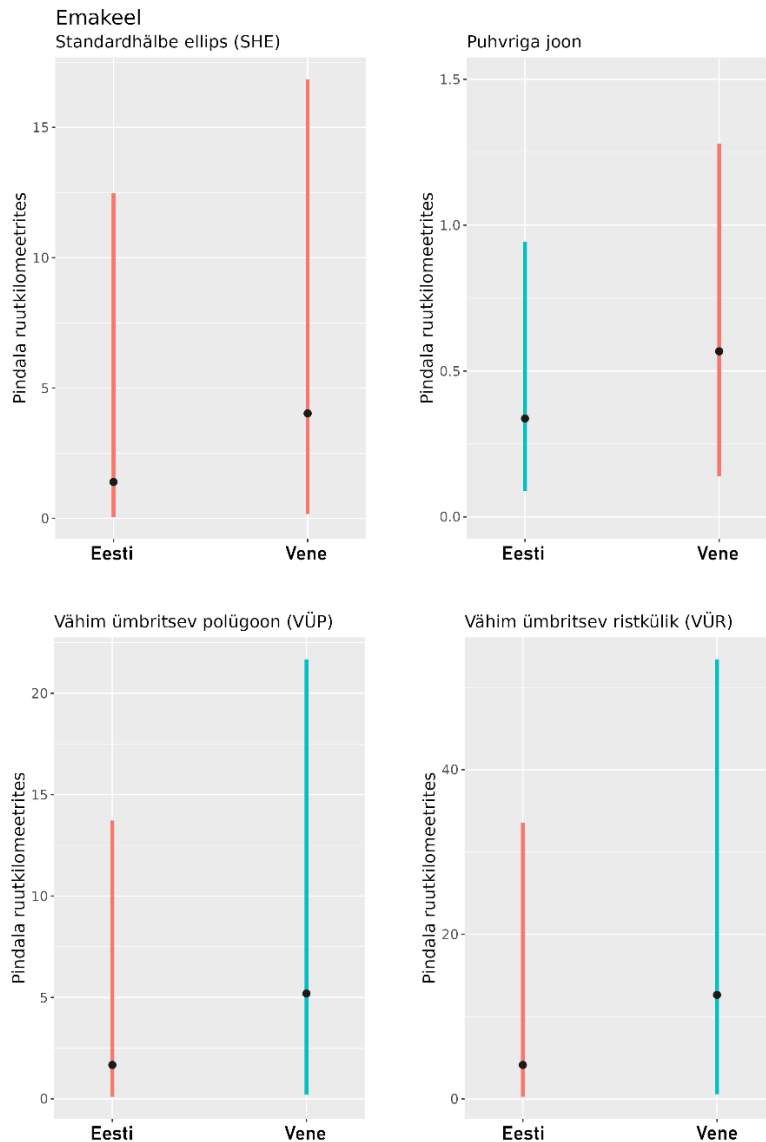
Kasutades sama testi puhul elukohta kontrollitava tunnusega leidis Duncani test kahe ruumikuju puhul statistiliselt olulise erinevuse (joonis 10). Puhvriga joon ning vähim ümbritsev ristkülik olid statistiliselt olulise erinevusega samas kui standardhälbe ellips ning vähim ümbritsev polügon määrati samasse gruppi.



Joonis 10. Duncani testi tulemused elukohapõhisel võrdlusel. Joon näitab kvartiilihaaret, punkt mediaanväärtust.

Kolmanda tunnuse, emakeele, põhisel võrdlemisel leidis Duncani test kõige rohkem statistiliselt erinevaid gruppe (joonis 11). Emakeelt võrdlusalusena kasutades olid statistiliselt olulise erinevusega puhvriga joon, vähim ümbritsev ristkülik ja vähim ümbritsev polügon.

Vaid standardhälbe ellipsi puhul ei leitud eesti ja vene keelt kõnelejate vahel statistiliselt olulist erinevust.



Joonis 11. Duncani testi tulemused kasutades võrdlusalusena emakeelt. Joon tähistab kvartiilihaaret ning punkt mediaanväärtust.

Duncani test tuvastas kõige rohkem erinevusi võrreldud gruppide vahel puhvriga joone ruumikuju puhul, kus kõigi kolme võrreldud tunnuse korral tuvastati statistiliselt oluline erinevus. Vähimal ümbritseval riskülikul tuvastati statistiliselt olulised erinevused kahe ning vähima ümbritseva polügoni puhul ühe tunnuse korral. Standardhälbe ellipsi puhul puudus kõigi kolme tunnuse puhul statistiliselt oluline erinevus.

Duncani testi kasutatakse tavaliselt kolme või enama grupi omavaheliseks võrdlemiseks. Selles töös analüüsiti algselt ka selliseid tunnuseid, kus oli kolm või enam gruppi, mistõttu võeti

kasutusele Duncani test. Töö käigus aga liikus fookus nendele kolmele tunnusele, kus on vaid kaks erinevat gruppi, kuid Duncani test jäi siiski kasutusse.

Duncani testi puhul on teada, et test kaldub tegema I-liiki vigasid ehk andma valepositiivseid tulemusi (Petrinovich & Hardyck, 1969). Selle võimaluse vähendamiseks ning vaid kahe võrdlusgrupiga tunnuste analüüsimise kompenseerimiseks seati Duncani testi olulisuse nivoo väärtuseks 0,01 ning kõikidele Duncani testiga analüüsitud paaridele tehti ka Kruskal-Wallise test.

Kruskal-Wallise test kontrollib, kas võrreldavad valimid ehk selle töö mõistes mingi tunnuse alusel grupeeritud ruumikujude pindalad on mediaanide võrdluse alusel üksteisest statistiliselt erinevad. Kruskal-Wallise test leidis, et kõikides analüüsitud paarides on tegemist statistiliselt oluliselt erinevate valimitega. Kruskal-Wallise testi tulemused on toodud tabelis 3.

Tabel 3. Kruskal-Wallise testi tulemused erinevate ruumikujude ja tunnuste põhjal.

Ruumikuju	Tunnus	Olulisuse tõenäosus
Vähim ümbritsev polügon	Elukoht	$p < 0,001$
Vähim ümbritsev ristkülik		$p < 0,001$
Standardhälbe ellips		$p < 0,001$
Puhvriga joon		$p < 0,001$
Vähim ümbritsev polügon	Emakeel	$p < 0,001$
Vähim ümbritsev ristkülik		$p < 0,001$

Ruumikuju	Tunnus	Olulisuse tõenäosus
Standardhälbe ellips	Emakeel	$p < 0,001$
Puhvriga joon		$p < 0,001$
Vähim ümbritsev polügon	Haridustase	$p < 0,001$
Vähim ümbritsev ristkülik		$p < 0,001$
Standardhälbe ellips		$p < 0,001$
Puhvriga joon		$p < 0,001$

4. Arutelu

4.1 Tulemuste analüüs

Korrelatsioonimaatriksi koostamise ja analüüsi tulemusena leidis kinnitust teooria, mille kohaselt on erinevad tegevusruumi kujutusviisid omavahel tugevas korrelatsioonis. Leiti tugevad seosed VÜPi, VÜRi, standardhälbe ellipsi ja puhvriga joone ruumikujude pindalade vahel, mis on kooskõlas Sila-Nowicka *et al* (2023) artikliga, mis samuti leidis, et nende näitajate vahel on väga tugev seos.

Ruumikujude pindalade mediaanide absoluutväärtuste võrdlemine tõi esile puhvriga joone ruumikuju kasutamisel puhvri raadiuse suure mõju absoluutväärtuste võrdlemisele ning seeläbi ka puhvri raadiuse valimise olulisust.

Duncani testi ja Kruskal-Wallise testil põhinevad tulemused näitavad, et elukoht, emakeel ning haridustase mõjutavad olulisel määral inimese tegevusruumi pindala. Olgugi, et Duncani testi tulemused viitasid puhvriga joonele, kui kõige kontrastsemale ruumikujule siis kindlaid järeldusi ei olnud selle testi põhjal võimalik teha. Kruskal-Wallise testi põhjal samuti ei joonistunud välja selgelt kõige kontrastsemat ruumikuju, küll aga kinnitas test, et erinevused kõikide võrreldud tunnuste alusel on statistiliselt olulised kõikide ruumikujude korral.

Ruumikujude pindalade mediaanväärtuste võrdluskordajate analüüs tõi väga selgelt esile kõige vähem kontrastse ruumikuju, milleks osutus puhvriga joon. Puhvriga joone võrdluskordaja oli kõikide tunnuste puhul ülejäänud meetoditest pea poole väiksem. Neljast võrreldud meetodist ülejäänud kolm olid sarnaste tulemustega. Seega võib analüüsi põhjal öelda, et kõige tõenäolisemalt on kolm sarnase võrdluskordajaga meetodit lähedal tegelikkusele ning erandliku väärtusega puhvriga joon esitab tulemusi kallutatult.

4.2 Soovitused tulevasteks uurimusteks

Uurimismeetodite loomumomaste mõjude mõistmine on kriitiline osa saadud tulemuste tõlgendamisest. Seetõttu on oluline jätkata kallutatuse uurimist nii selle meetodi kui ka teiste inimgeograafias kasutatavate meetodite puhul.

Tulenevalt selle töö leidudest soovib töö autor edasistes uurimustes pöörata rohkem tähelepanu GPS-andmete põhjal arvatud ruumikujude loomulikule kalduvusele ning võtta seda aspekti ruumikujude valimisel arvesse.

Seoses GPS-andmete suure andmemahu ning seetõttu ka ruumikujude arvutamiseks ja nendega tehete tegemiseks kuluva energiakuluga soovib töö autor ka seda aspekti ruumikujude valimisel arvestada. Juhul kui kõik muud tegurid on võrdsed soovib töö autor eelistada

ruumikujude kasutamisel vähimat ümbritsevat riskülikut, sest tegemist on kõige väiksemat andmemahdu nõudva ruumikujuga ning võimalusel vältida puhvriga joone kasutamist, sest analüüsitud ruumikujudest oli just see kõige suurema andmemahuga. Tegemist on olulise viisiga, kuidas panustada teadustööde keskkonnasäästlikkusse, sest olgugi, et ühe ruumikuju hoiustamiseks vajalik andmemahd on küllaltki väike, siis tuhandete ruumikujude korral muutub andmemahu hulk kiiresti väga suureks.

5. Kokkuvõte

Suurandmetel põhinevat teadust tehes on lihtne usaldada väljatöötatud algoritme ning mudeleid, sest tihti ei mõista inimesed enam kõike, mida masinad keerulisi arvutusi tehes teevad. Sellest tulenevalt võib ka geograafias esineda olukordi, kus algoritmi poolt leitud tulemused võivad olla kallutatud ning teadlane ise ei ole sellest teadlik.

Selle töö eesmärgiks on vältida sellise olukorra teket inimese tegevusruumiuuringutes. Selleks uuriti Tallinna Kalamaja ning Priisle elanike käest kogutud GPS-andmete põhjal arvatud tegevusruumide ruumikujusid. Andmete kogumiseks kasutati mobiilirakendust MobilityLog ning selle töö raames kasutatud andmete kogumine toimus ajavahemikus 19.10.2020 kuni 18.10.2021. Kokku kasutati andmeanalüüsis veidi vähem kui 26 000 andmerida ühe ruumikuju kohta.

Leidmaks, kas mõnel ruumikujul esineb iseloomulik kallutus võrreldi omavahel kõikide ruumikujude pindalade mediaanväärtusi kasutades selleks nii absoluutväärtuste võrdlemist, võrdluskordajat, Duncani testi kui ka Kruskal-Wallise testi.

Analüüsi tulemusena selgus, et kõikide uuritud tunnuste puhul on absoluutväärtusena kõige suurema gruppide vahelise erinevusega vähim ümbritsev ristkülik, millele järgnevad vähim ümbritsev polügon ning standardhälbe ellips. Kõige väiksema absoluutväärtuste erinevusega on puhvriga joone ruumikuju.

Analüüsidest võrdluskordajate väärtuste analüüsil ilmnnes samuti, et puhvriga joon kujutab gruppide vahelisi erinevusi teistest meetoditest pea poole väiksemana ehk puhvriga joon näitas võrdlusgruppe omavahel sarnasemana kui vähim ümbritsev polügon, vähim ümbritseb ristkülik ning standardhälbe ellips, mis kõik näitasid gruppide vahelisi erinevusi peaaegu sama suurena. Seetõttu saab öelda, et iseloomulik kallutus esineb kõige tõenäolisemalt puhvriga joone ruumikujul.

Selleks, et kontrollida, kas saadud tulemused on ka statistiliselt olulised ning selleks, et neid paremini visualiseerida viidi läbi ka Duncani test ning Kruskal-Wallise test. Nende kahe testi tulemusena võib kinnitada, et kõikide analüüsitud tunnuste alusel ning kõikide ruumikujude puhul esines statistiliselt olulisi erinevusi gruppide vahel.

Antud töö tulemuste põhjal saab öelda, et analüüsitud tegevusruumi kujutamise meetoditest ühe puhul esineb iseloomulik kallutus, millega tuleks teadustöö metoodika valimisel arvestada. Samuti saab öelda, et andmemahu kokkuhoidmise eesmärgil on kõige kasulikum kasutada vähimat ümbritsevat ristkülikut, sest see näitab samasuguseid tulemusi vähima

ümbritseva polügoni ja standardhälbe ellipsiga, kuid nõuab hoiustamisel vähem andmemahtu ning on seetõttu ka arvutuste tegemisel energiasäästlikum. Samal põhjusel tuleks võimalusel vältida puhvriga joone kasutamist, sest tegemist on ülimalt mahuka ruumikuuga.

A comparative analysis of GPS-data based methods used for representing activity spaces

Jürgen Öövel

Summary

As natural sciences have shifted towards using more big data in research a new problem has arisen. Researchers rely more on complex algorithms and systems that can be too complex to understand leading to the so-called black box problem, where people blindly trust the results that a program or algorithm outputs.

In a broader sense, this phenomenon can occur in geography as well, when researchers working with big data just use algorithms that have been used before without fully understanding why. Therefore, it is critical to understand the inherent biases of the methods that scientists use in their research.

The aim of this thesis is to identify if any of the spatial representations of human activity spaces have built in biases. This thesis investigates four different spatial representations of activity space: minimum convex polygon, standard deviation ellipse, bounding box and daily path that has a buffer with 20m radius surrounding it.

All four methods are calculated based on the GPS data collected using the Android phone app MobilityLog. This data was collected from 184 residents of Kalamaja and Priisle areas in Tallinn, Estonia as a part of the research project “Understanding the Vicious Circles of Segregation. A Geographic Perspective”. Data used in this thesis was collected between 19 October 2020 and 18 October 2021.

First part of the analysis was the creation of a Spearman correlation matrix of different calculated attributes of the activity space shapes. This was done to better understand the dataset and to find correlations between different parameters of the four methods. The second part of the analysis was to compare the median values of the areas of the different methods. The reason for choosing the median as the representation of the average value is because the dataset includes a lot of abnormally high values meaning that the dataset has a non-normal distribution which was confirmed using the Lilliefors test and the Pearson's chi-squared test.

Comparison of median values showed that the bounding box has the biggest differences and the daily path area has the smallest differences when comparing the median area sizes of different groups. Because the daily path area had significantly smaller areas than the rest of the methods, a new value was introduced- the comparison factor. Comparison factor is a value that

shows how many times larger is the larger median area compared to the smaller median area of a comparison pair. Analyses of the comparison factors also supported the previous findings and showed that the daily path area showed almost two times smaller differences between comparison groups when compared to the rest of the methods.

In order to confirm that the findings are statistically significant and to better visualise the results the Duncan test and Kruskal-Wallis test were performed. Results of both tests confirmed the statistical importance of the differences between all groups with all the methods used for describing activity spaces.

Results of the data analysis show that daily path area has built in bias. This bias must be considered when using this method for representing the daily activity space in research. Furthermore, the author suggests that bounding box should be the preferred method for calculating activity spaces. This is because this method needs the least amount of space per shape to be stored and therefore uses the least amount of computing power when used in calculations, while providing similar results to standard deviation ellipse and minimum convex polygon. On the other hand, daily path area should be avoided if possible as it requires a lot more storage space.

As this thesis only included data collected in a relatively small area and from a relatively small amount of people, the author suggests that further research is needed to even better understand the biases of the methods used for calculating activity spaces. Furthermore, the author suggests that this kind of bias research should be expanded to other methods used in human geography as well.

Tänuavaldused

Soovin südamest tänada oma juhendajat Anto Aasat, kes oli töö koostamisel mulle igati toeks ning tutvustas mulle Ri kasutamise maagiat. Samuti soovin tänada Martin Haamerit ingliskeelsetele mõistetele vastete ning tõlgete pakkumise eest ning Mirian Valki statistiliste testide lahtimõtestamise eest.

Ühtlasi soovin tänada enda perekonda toetava ning mõistva suhtumise eest, sõpru Aliiset ja Adelet motiveerimise ja moraalse toetuse eest ning enda Kaitseväge ülemaid vastutulelikkuse eest.

Soovin tänada ka Rotalia Foundationit ning Tartu Ülikooli Sihtasutust minu õpingute toetamise eest.

Jürgen Öövel

23.05.2024

Kasutatud kirjandus

- Ahas, R., Aasa, A., Silm, S., & Tiru, M. (2010). Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data. *Transportation Research Part C: Emerging Technologies*, 18(1), 45–54. <https://doi.org/10.1016/j.trc.2009.04.011>
- Ahas, R., Aasa, A., Yuan, Y., Raubal, M., Smoreda, Z., Liu, Y., Ziemlicki, C., Tiru, M., & Zook, M. (2015). Everyday space–time geographies: Using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn. *International Journal of Geographical Information Science*, 29(11), 2017–2039. <https://doi.org/10.1080/13658816.2015.1063151>
- Brożek, B., Furman, M., Jakubiec, M., & Kucharzyk, B. (2023). The black box problem revisited. Real and imaginary challenges for automated legal decision making. *Artificial Intelligence and Law*. <https://doi.org/10.1007/s10506-023-09356-9>
- Etis. (i.a.). *Ruumilise eraldatuse nõiaringi geograafiline analüüs*. <https://www.etis.ee/Portal/Projects/Display/7f19f028-50b5-4b79-a792-c0e04bfe791b> (viimati vaadatud 23.05.2024).
- Eurostat. (2014). *Feasibility study on the use of mobile positioning data for tourism statistics :consolidated report*. Publications Office. <https://data.europa.eu/doi/10.2785/55051>
- Gavela, E. S. O. (2020). *Comparing activity-space based segregation methods: A study with GPS data* [Magistritöö, Tartu Ülikool]. <http://hdl.handle.net/10062/68397>
- Haamer, M. (2022). *Tallinna ühistranspordi ligipäätavuse hindamine GPS-andmete põhjal* [Magistritöö, Tartu Ülikool]. <http://hdl.handle.net/10062/82867>
- Hägerstrand, T. (1970). What about people in Regional Science? *Papers of the Regional Science Association*, 24(1), 6–21. <https://doi.org/10.1007/BF01936872>

- Hirsch, J. A., Winters, M., Clarke, P., & McKay, H. (2014). Generating GPS activity spaces that shed light upon the mobility habits of older adults: A descriptive analysis. *International Journal of Health Geographics*, 13(1), 51. <https://doi.org/10.1186/1476-072X-13-51>
- IEA. (i.a.). *Data centres & networks*. IEA. <https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks>. (viimati vaadatud 23.05.2024).
- IMO. (2023). *Nutitelefonid rakenduse MobilityLog andmed*. <https://imo.ut.ee/taristu/nutitelefonid-rakenduse-mobilitylog-andmed/> (viimati vaadatud 23.05.2024).
- Jankowska, M. M., Yang, J.-A., Luo, N., Spoon, C., & Benmarhnia, T. (2023). Accounting for space, time, and behavior using GPS derived dynamic measures of environmental exposure. *Health & Place*, 79, 102706. <https://doi.org/10.1016/j.healthplace.2021.102706>
- Järv, O., Ahas, R., & Witlox, F. (2014). Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. *Transportation Research Part C: Emerging Technologies*, 38, 122–135. <https://doi.org/10.1016/j.trc.2013.11.003>
- Kozlovskiy, N. (2023). *The influence of socio-demographic factors and place of residence on people's activity space: A study based on GPS data in Tallinn, Estonia* [Magistritöö, Tartu Ülikool]. <https://hdl.handle.net/10062/90698>
- Kurvits, K. (2020). *Ööpäevase sotsiaalse rütmi piirkondlikud erinevused Eestis* [Bakalaureusetöö, Tartu Ülikool]. <https://dspace.ut.ee/handle/10062/68344>
- Linde, E. (2016). *Tallinna ajakasutuse rütmid* [Bakalaureusetöö, Tartu Ülikool]. <http://hdl.handle.net/10062/52956>

Maa-amet. (2024). *Haldus- ja asustusjaotus.*

<https://geoportaal.maaamet.ee/est/ruumiandmed/haldus-ja-asustusjaotus-p119.html>

(viimati vaadatud 23.05.2024).

Millward, H., Hafezi, M. H., & Daisy, N. S. (2019). Activity travel of population segments grouped by daily time-use: GPS tracking in Halifax, Canada. *Travel Behaviour and Society*, 16, 161–170. <https://doi.org/10.1016/j.tbs.2019.05.005>

Neutens, T., Witlox, F., & Demaeyer, P. (2007). Individual accessibility and travel possibilities: A literature review on time geography. *European Journal of Transport and Infrastructure Research*. <https://doi.org/10.18757/EJTIR.2007.7.4.3403>

Newsome, T. H., Walcott, W. A., & Smith, P. D. (1998). Urban activity spaces: Illustrations and application of a conceptual model for integrating the time and space dimensions. *Transportation*, 25, 357–377. <https://doi.org/10.1023/A:1005082827030>

Paris, M. (2022). *Etnilisuse mõju tegevuskohtade külastatavusele väljaspool Tallinna* [Magistritöö, University of Tartu]. <http://hdl.handle.net/10062/82870>

Petrinovich, L. F., & Hardyck, C. D. (1969). Error rates for multiple comparison methods: Some evidence concerning the frequency of erroneous conclusions. *Psychological Bulletin*, 71(1), 43–54. <https://doi.org/10.1037/h0026861>

Polsby, D. D., & Popper, R. (1991). The Third Criterion: Compactness as a Procedural Safeguard Against Partisan Gerrymandering. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2936284>

Poom, A. (2019). Tegevusruumiuringud GPSi andmete abil: Mobiilsusuuringute labori andmeinfrastruktuur. *Uurimusi eestikeelse geograafia 100. Aastapäeval. Publicationes Instituti Geographici Universitatis Tartuensis 113* (pp 204–214). Tartu Ülikool.

Raun, J. (2020). *Mobile positioning data for tourism destination studies and statistics* [Doktoritöö, Tartu Ülikool]. <http://hdl.handle.net/10062/67430>

- Roosaare, J., Mõisja, K., Aunap, R., Oja, T., & Sepp, E. (2019). *Geoinformaatika*. Tartu Ülikooli Kirjastus.
- Saluveer, E., Raun, J., Tiru, M., Altin, L., Kroon, J., Snitsarenko, T., Aasa, A., & Silm, S. (2020). Methodological framework for producing national tourism statistics from mobile positioning data. *Annals of Tourism Research*, 81, 102895. <https://doi.org/10.1016/j.annals.2020.102895>
- Schönfelder, S., & Axhausen, K. W. (2003). Activity spaces: Measures of social exclusion? *Transport Policy*, 10(4), 273–286. <https://doi.org/10.1016/j.tranpol.2003.07.002>
- Sheller, M., & Urry, J. (2006). The New Mobilities Paradigm. *Environment and Planning A: Economy and Space*, 38(2), 207–226. <https://doi.org/10.1068/a37268>
- Sila-Nowicka, K., Fotheringham, A. S., & Demšar, U. (2023). Activity triangles: A new approach to measure activity spaces. *Journal of Geographical Systems*, 25(4), 489–517. <https://doi.org/10.1007/s10109-023-00414-z>
- Smith, L., Foley, L., & Panter, J. (2019). Activity spaces in studies of the environment and physical activity: A review and synthesis of implications for causality. *Health & Place*, 58, 102113. <https://doi.org/10.1016/j.healthplace.2019.04.003>
- Statistikaamet. (2018). *Statistikaameti arengukava 2018-2022*. Kättesaadav: <https://www.stat.ee/et/statistikaamet/meist/strateegia> (viimati vaadatud 23.05.2024)
- Tallinna Linnaplaneerimise Amet. (2020). *Ruumiandmed / Tallinn*. <https://www.tallinn.ee/et/geoportaal/ruumiandmed> (viimati vaadatud 23.05.2024).
- Tominga, A., Silm, S., Orru, K., Vent, K., Klaos, M., Võik, E.-J., & Saluveer, E. (2023). Mobile positioning-based population statistics in crisis management: An Estonian case study. *International Journal of Disaster Risk Reduction*, 96, 103887. <https://doi.org/10.1016/j.ijdr.2023.103887>

Zernant, K. H. (2023). *Looduslike alade külastamine Tallinna elanike seas mobiiltelefoni GPS andmetel* [Magistritöö, University of Tartu]. <https://hdl.handle.net/10062/90750>

Lisa 1. Mediaanväärtuste võrdlustabelid

Tabel 4. Erinevate ruumikujude pindalade mediaanväärtuste võrdlus. Tumedalt on märgitud võrdluspaari suurem väärtus.

Ruumikuju	Tunnus	Tunnuse väärtus	Tegevusruumi pindala mediaanväärtus (km ²)	Mediaanväärtuste erinevus (km ²)	Võrdluskordaja
Puhvriga joon	Emakeel	Eesti	0,3372248	0,2304227	1,683291087
		Vene	0,5676475		
VÜR	Emakeel	Eesti	4,12175	8,51314	3,065418815
		Vene	12,63489		
VÜP	Emakeel	Eesti	1,671116	3,521278	3,107141575
		Vene	5,192394		
Standardhälbe ellips	Emakeel	Eesti	1,393695	2,63446	2,890270109
		Vene	4,028155		
Puhvriga joon	Haridus	Kõrgem	0,3704838	0,2449667	1,661207589
		Madalam	0,6154505		
VÜR	Haridus	Kõrgem	5,08468	8,81814	2,73425663
		Madalam	13,90282		
VÜP	Haridus	Kõrgem	2,153867	3,214917	2,492625589
		Madalam	5,368784		
Standardhälbe ellips	Haridus	Kõrgem	1,78748	2,360145	2,320375613
		Madalam	4,147625		
Puhvriga joon	Elukoht	Kalamaja	0,2931393	0,2977205	2,015628065
		Priisle	0,5908598		
VÜR	Elukoht	Kalamaja	3,154252	10,058211	4,188778512
		Priisle	13,212463		

VÜP	Elukoht	Kalamaja	1,371231	3,753189	3,737094625
		Priisle	5,12442		
Standardhälbe ellips	Elukoht	Kalamaja	1,15119	2,788997	3,422707807
		Priisle	3,940187		

Tabel 5. Erinevate ruumikujude ümbermõõtude mediaanide väärtused. Tumedalt on märgitud võrdluspaari suurem väärtus.

Ruumikuju	Tunnus	Tunnuse väärtus	Tegevusruumi pindala mediaanväärtus (km ²)	Mediaanväärtuste erinevus (km ²)	Võrdluskordaja
Puhvriga joon	Emakeel	Eesti	14,47419	9,36092	1,64673
		Vene	23,83511		
VÜR	Emakeel	Eesti	8,964521	7,25342	1,809125
		Vene	16,217941		
VÜP	Emakeel	Eesti	6,903581	6,060037	1,877811
		Vene	12,963618		
Puhvriga joon	Haridustase	Kõrgem	15,58187	10,64354	1,683072
		Madalam	26,22541		
VÜR	Haridustase	Kõrgem	9,886382	8,04972	1,814223
		Madalam	17,936102		
VÜP	Haridustase	Kõrgem	7,658055	6,558478	1,856416
		Madalam	14,216533		
Puhvriga joon	Elukoht	Kalamaja	12,40827	3,214917	2,029743
		Priisle	25,1856		
VÜR	Elukoht	Kalamaja	7,672353	2,360145	2,241864
		Priisle	17,200375		

VÜP	Elukoht	Kalamaja	5,963481	7,996112	2,340846
		Priisle	13,959593		

Lihtlitsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, **Jürgen Öövel**,

(autori nimi)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose
GPS-andmete põhjal arvutatud tegevusruumide kujutamise meetodite võrdlev analüüs,
(lõputöö pealkiri)

mille juhendaja on **Anto Aasa**,

(juhendaja nimi)

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Jürgen Öövel

23.05.2024