Master's thesis in Geography

Geoinformatics

Capturing International and Domestic Mobility Patterns of Minority Language Groups: Case of Finland using Twitter data

Emil Mattias Ehnström

2021

Supervisors:
Olle Järv
Tuuli Toivonen

Master's Programme in Geography

Faculty of Science

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Faculty	Department
Faculty of Science	Department of Geosciences and Geography

Author

Emil Mattias Ehnström

Title of thesis

Capturing International and Domestic Mobility Patterns of Minority Language Groups: Case of Finland using Twitter data

Programme and study track

Master's programme in geography, Geoinformatics

Level of the thesis	Date	Number of pages
Master's thesis, 30 credits	September 2021	63 + 2 appendixes

Abstract

The number of people belonging to a language minority in Finland is increasing and people are becoming more and more spatially mobile. This has also led to an increase in transnationals and higher rates of cross-border mobility. With new methods involving social media big data, we can map spatial mobility patterns in new ways and deepen the understanding of how people relate to space. Differences in spatial mobility can for example give us an indication of the rate of integration into society. Some claim that a more spatially mobile life is a sign of success, but can we see differences in spatial mobility between people in Finland?

The three language minorities considered in this thesis are Swedish, Russian, and Estonian. The history and culture of these groups are different as well as their status in Finnish society. Swedish speakers, with a national language status, have a different role in society, but do this well integrated minority differ from the other ones spatially?

By using Twitter data and looking at the spatial mobility within Finland, we see where differences occur between language groups. To understand how strong ties the language groups have with neighbouring countries, we look at cross-border mobility to Estonia, Russia, and Sweden.

The results show that there are differences in the spatial mobility of language minorities in Finland. Estonian speakers most frequently visit Estonia, while at the same time they are less mobile within Finland. The variation was large for Russian speakers, with some visiting Russia often and others almost never. Swedish speakers seem to have relatively weak ties to Sweden, compared to the other language groups and have very similar spatial mobility to the majority Finnish speaking population.

Keywords

Cross-border mobility, Twitter, language minority, social media,

Where deposited

HELDA

Additional information

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Fakultet	Institution
Matematisk-naturvetenskapliga fakulteten	Avdelningen för geovetenskaper och geografi

Författare

Emil Mattias Ehnström

Avhandlingens titel

Språkminoriteters gränsöverskridande-och inrikesrörlighet i Finland – Twitterdataanalys

Utbildningsprogram och studieinriktning

Magisterprogrammet i geografi, Geoinformatik

Avhandlingens nivå	Datum	Sidantal
Pro Gradu, 30 studiepoäng	September 2021	63 + 2 bilagor

Referat

Antalet människor som tillhör språkminoriteter ökar i Finland och överlag blir människor alltmer rumsligt mobila. I takt med detta har även den gränsöverskridande rörligheten samt antalet transnationella personer ökat. Med nya metoder som involverar stora mängder data från sociala medier, så kan vi kartlägga rumsliga rörelsemönster på nya sätt och öka vår förståelse för människors plats i rummet. Skillnader i rumslig rörlighet kan till exempel indikera integrationsgraden i samhället. Vissa argumenterar för att en hög rumslig rörlighet är en måttstock för ett lyckat liv, men kan vi hitta skillnader i den rumsliga rörligheten mellan folkgrupper i Finland?

Språkminoriteterna som behandlas I denna avhandling är personer som talar svenska, ryska samt estniska. Deras historia och kultur samt status i det finska samhället skiljer sig märkbart från varandra. Svenskan är nationalspråk I Finland, vilket skapar en annan roll i samhället för de svenskspråkiga, men syns den här skillnaden i rumsligt beteende?

Genom att se på den rumsliga rörligheten inom Finland med hjälp av Twitterdata, så kan vi iaktta skillnader mellan språkgrupperna. För att förstå hur starka band språkgrupperna har med grannländerna, tittar vi även på den gränsöverskridande rörligheten till Estland, Ryssland och Sverige.

Resultaten visar att det finns skillnader i rumsliga rörelsemönster mellan språkminoriteter i Finland. Personer som talar estniska besöker Estland ofta, samtidigt som deras rumsliga rörelser inom Finland är lägst i undersökningen. Variation I den rumsliga rörligheten var störst bland ryskspråkiga, vissa besökte Ryssland ofta medan andra väldigt sällan. Svenskspråkiga verkar ha relativt svaga band till Sverige I jämförelse med de andra språkminoriteterna och deras rumsliga rörligheten är väldigt lik den finskspråkiga majoritetsbefolkningen.

Nyckelord

Gränsöverskridande rörlighet, Twitter, språkminoritet, sociala medier

Förvaringsställe

HELDA

Övriga uppgifter

Table of Contents

Table of Figures

Table of Tables

1 Introduction

Imagine a beach abroad. You take out your smartphone and take a picture of this paradise on earth. Then you think to yourself that this is a picture worth sharing. So, you do. To make sure your followers know where you are you also include the location as a tag. Time flies and you forget all about posting that image, never thinking that the picture could be part of scientific research one day.

Social media posts contain plenty of information and one important one in particular is the geotag. This tag reveals the location of the social media post. When an immense amount of people uses these geotags it turns into what we would call big data. Then the metadata, containing the location, can be analysed for example to discover mobility patterns of people. These patterns can be very helpful to identify and understand how our society functions from various perspectives.

Humans are mobile beings. We are constantly moving around both in space and in society. *Human mobility* explains our way of moving through space (Kellerman, 2012). The human society is also mobile, containing people, information, and goods. Human society is also experiencing an increase in movements, the further society develops, the more mobile we seem to be. This in turn has led to social sciences focusing on human mobility through the *New Mobilities Paradigm* (Sheller & Urry, 2006). Where the focus lies on an individual's movements and its relationship to spaces.

Movements happen everywhere, and sometimes they happen across borders. Borders are something that define and divide space. Border can take different shapes, some are physical, visible in the landscape, whereas others based on social or cultural factors (Lynch, 1960; Minghi, 1963).

Individuals move within spaces but also across them, which is referred to as *cross-border mobility*. People move across borders for different reasons, it can be in the form of migration, where the usual residency is changed (International Organization for Migration, 2021). Reasons for cross-border mobility can be work or education (Mazzoni et al., 2017; Nonnenmacher et al., 2021). When people of different backgrounds move to

a new place, it is often assumed that the longer a person stays in the country the more integrated they become and are then less likely to return to their place of origin (de Haas & Fokkema, 2011; Gordon, 1964). This traditional view has been contested, for example by Anniste & Tammaru (2014) that found that the well-integrated Estonians in Finland were more likely to return to Estonia instead of the less integrated Russian minority. Integration and segregation research has shifted towards a person-based approach and together with positional data it is possible to determine dynamics of ethnic segregation and socio-economic factors.

In this thesis I am interested in the three largest language groups in Finland after Finnish, Swedish, Russian, and Estonian. Historically and culturally these language groups differ from each other (Daher et al., 2016). The population structure in Finland is changing and minorities are increasing (Official Statistics of Finland (OSF), 2021), therefore it is important to understand their role in society. My hypothesis is that these three language groups also have different spatial mobilities, that in turn can help explain the rate of integration into Finnish society. A low spatial mobility within a country might indicate that the group is less integrated in society (Järv et al., 2015). I hope to find if low spatial mobility within Finland is correlating with the transnational ties to a majority language country. Strong ties to another country than Finland, might indicate a transnational identity (Silm et al., 2020).

The main research questions of this thesis are:

1. How are people, belonging to a language minority in Finland spatially connected with neighbouring countries?
    - How often do people visit neighbouring countries where the majority language is their native language?
    - How diverse are the mobility patterns for the language groups?
2. How are people, belonging to a language minority in Finland spatially connected within Finland?
3. How do the spatial mobility patterns of the language groups differ?

As a method for exploring these spatial issues I turn to geo-located Twitter data. By identifying the language, a user uses I can aggregate users into language groups. These

groups can act as a proxy for actual populations and provide insights on mobility patterns (Hawelka et al., 2014). With visual inspection, diversity indices and statistical analyses I try to gain a deeper understanding of the differences in spatial mobility between language groups in Finland.

2 Background

2.1 Human Mobility

In the most fundamental sense, *mobility* is, the "quality or state of being mobile or movable" (Merriam-Webster Inc., n.d.-a). In general terms, this can be applied for various entities, such as goods, services, groups, and individuals. When specifying *human mobility*, it could be considered as the "ability of the human body to move across space" (Kellerman, 2012, p. 1).

The *new mobilities paradigm* or mobility turn has become one of the central ideas in current mobility research, it has even gone so far that is has been suggested to be "the new normal" (Sheller & Urry, 2016, p. 21) within social sciences. Traditionally social sciences research has been done from a sedentary perspective, thinking of the issue in a sort of closed space as well as within territorial boundaries (Sheller & Urry, 2006). In the new mobilities paradigm, the focus includes the relationships and connections with a space as well as a deterritorialization of spatial boundaries.

Another emphasis lies on spatial mobility in multiple forms, expanding our way of thinking about human mobility (Sheller & Urry, 2006). Our networks are not confined and the interdependence as well as connection between all forms of mobility are important. Places in themselves are not seen as static but dynamic (Sheller & Urry, 2006). A place is compared to a ship, that can move around, not remaining in the same place over time. A key aspect of places are the relationships it entails. This could be for instance the relationship between and individual and the place.

(Faist, 2013) requests a more critical viewpoint of the new mobilities paradigm. He asks some important questions: "Is mobility really a human universal, as anthropologists tell us? And "is it true that spatial mobility is a marker of success in navigating the global world?" (Faist, 2013, p. 1644).

2.1.1 Spatial Mobility

Spatial mobility as (Kaufmann et al., 2004) so well puts it "refers to geographic displacement, i.e., the movement of entities from an origin to a destination along a specific trajectory that can be described in terms of space and time". The entity could for instance be an individual, goods or a group of people. An important theory that Hägerstrand (1970) developed is the time-space concept, or time geography, where he emphasizes the individual and how there are not only space coordinates but also time coordinates to be considered. An entity does not only exist in space, but it also exists in time simultaneously. He introduces the *life path,* which is the path an individual takes from birth to death. The premise is that the aim of an individual is to survive. During a person's life path, it is not possible to skip time, one is always tied to a certain time on the time scale. During this voyage through life a person has certain constraints that affect the freedom of the life path. Hägerstrand mentions that there are a high number of different constraints but that most of them can be divided the into three families: *Capability constraints*, *coupling constraints*, and *authority constraints*.

In the first constraint category we have an individual's biological constraints. This could be for example sleeping and eating, that at some point in time becomes unavoidable. Since these are a necessity the limit the time that could be used in other ways. Hägerstrand talks about *tubes* that can be viewed as symbols for accessible space. The inner tube is what a person can reach with its own body whereas the second tube is how far a person could be reach through oral and visual communication. The third tube is linked to a person's home and so forth. Most people require a place to stay for resting and keeping belongings. This place affects what space that is accessible for a person. This space becomes a person's *island*. The size of the island is affected by the available transportation modes. Hägerstrand visualises the island as one or multiple prisms on the day path (a small part of a life path), where under no circumstances a person can be outside of the prism.

Inside prisms we find coupling constraints. These constraints are a posed by the interaction with others. Most workplaces are only operating during a certain time and depending on set schedules the individual might be bound to be at a certain place at a certain time. Another example is schools where they have a strict timetable on when and when not to

be there. The coupling constraints are also connected to the home base, since they determine the space of the prism, together with the opportunities of transport.

The third and last category of constraints are linked to authority. In all societies there are certain power structures that can limit a person's space. On a macro level there are differences in what space you can access depending on for example your passport. Some places might be accessible after getting a visa and others might be totally off limits. Depending on your power status you have access to different spaces. On a microscale an authority constraint could be for example grandpa's favourite chair that is not accessible, at all times, for everyone.

Hägerstrand's thoughts remain relevant as a conceptual framework for understanding time geography, but it has been challenging to implement as a proper analytical framework (Miller, 2005). The original time-space concept encompasses physical space, but information and communications technology, ICT, has developed since then and we now also speak of a *virtual space*, stretching our understanding of space to also include digitally created spaces. (Shaw & Yu, 2009) points out that the original time geography concepts in a physical space are not possible to implement in virtual space, due to different characteristics of virtual space. Modern ICT allows us to be in touch with each other around the globe, questioning the importance of both temporality and spatiality (Ma, 2011). The original constraints have a change of character due to the rise of ICTs.

2.1.2 Social Mobility

The term *social mobility* is widely used in sociology, but it is also connected to spatial mobility and geography (Kaufmann et al., 2004). The concept suggests that there is room for movement in the social strata (Sorokin, 1941). Changing social status can be related to the spatial mobility, in other words a shift to a higher social class might also lead to an increase in spatial mobility (Kellerman, 2012). With a higher social status, one might have access to places that would otherwise be off limits due to for example Hägerstrand's (1970) coupling and authority constraints (Kaufmann et al., 2004; Kellerman, 2012).

The concept of *motility* aims to combine spatial and social mobility into a broader theory not only focusing on the past and present mobility, but also on the potential mobility

(Kaufmann, 2003; Kaufmann et al., 2004). The motility idea sees mobility as a form of capital. Depending on different factors, an entity (for example a person) has a certain capacity of being mobile, therefore the idea of "Motility: Mobility as Capital" (Kaufmann et al., 2004, p. 1). The three central elements in motility are *access*, *competence*, and *appropriation* (Kaufmann et al., 2004).

Access in motility refers to the accessible mobilities present in that moment (Kaufmann et al., 2004). This incorporates all the potential places reached in all available transport modes and the access can vary depending on the competencies. There are different forms of competence that Kaufmann et al. (2004) mentions: Physical ability, acquired skills and organizational skills. The first one mentioned includes the abilities of physical movements (for example how far can you walk during a certain time). Acquired skills is something that might be restricted due to legislation or rules, such as having a visa for entering a certain area. In this thesis, we can think of the differences in travel from Finland to Russia compared to other neighbouring countries. In general, a Finnish resident can travel to all neighbouring countries without a visa, except Russia, leading to a higher threshold for visiting the country.

Organizational skills are referring to abilities of gathering information and planning skills, such as knowing how to access a certain place through a road network(Kaufmann et al., 2004). The appropriation indicates how an entity chooses to interact with the access and competencies available. This means that all three central elements are closely linked to each other. In more recent years there has been an effort to typify this theory and applying it to an actual situation (Kaufmann et al., 2018). With a typology of motility, it is easier to describe the various skills that are needed to be mobile. Motility can be used to understand the complexity of mobility in relation to different social variables.

2.2 Activity Spaces

One way to measure the mobility and practices of individuals is through the concept of *activity spaces* (Farber et al., 2012; Hasanzadeh et al., 2019; Järv et al., 2014; Schönfelder & Axhausen, 2003; Shaw & Yu, 2009; Wong et al., 2011). The idea of activity spaces is closely linked to the concept of action spaces, that describes the total interactions and responses for an individual with its environment (Golledge & Stimson, 1997; Jakle et al.,

1976). Activity space is the part of an action space that focuses on the movement. (Golledge & Stimson, 1997) emphasises three aspects of an individual's activity space: The movements close to and at home of the individual, the movements to and from work, shops, hobbies and so on, considered regular activity locations and lastly the movements within and close by these regular activity locations. Another aspect Golledge & Stimson (1997) discusses, is the purpose of the trips. For work and social trips, it's obvious, and for other activity locations it can vary. However, the trips in themselves might not only have one purpose, but they might also be multipurpose trips that combine several things.

A person's perception of space can determine whether he or she will incorporate it in their activity space (Golledge & Stimson, 1997). If a person would be robbed on a certain street at a certain time, it might avoid that place in the future, changing the extent of his or her activity space. Another important part shaping a person's activity space is their role in society and their preferences and habits (Golledge & Stimson, 1997; Horton & Reynolds, 1970). Different mobility patterns are linked to the behaviour a person has, and this can change over time, giving activity space a temporal dimension (Järv et al., 2014; Silm & Ahas, 2014). A child can have a very different activity space than a retiree. It's worth noticing how social factors such as age, gender or ethnic background can change the way we use physical space (Järv et al., 2014; Shirazi, 2018; Silm et al., 2017).

2.3 Segregation & Inequality

Different social factors, like ethnicity, is also related to how segregated or spatially and socially isolated, groups become (Vanderbeck, 2006). It also appears on different scales. Segregation is multidimensional issue including a spatial, social, and temporal sphere (Järv et al., 2015; Massey & Denton, 1988). Advances in technology and access to big data has brought in new methods of studying segregation through for example mobile phone data (Järv et al., 2015; Toomet et al., 2015). Mostly research on segregation using the activity space approach has focused on the spatial scale within a single city or country (Farber et al., 2012; Hasanzadeh et al., 2019; Järv et al., 2015; Wong et al., 2011; Zhang et al., 2019) but in recent years there has been an urge to study the concept through a multi-country view to deepen the understanding of inequalities in a global world (Mooses et al., 2020). A higher social class in most cases indicate a higher transnational mobility

(Delhey et al., 2015), but for example Feng & Page (2010) found that the Chinese minority in New Zealand is more transnationally mobile than the majority population and Mooses et al. (2020) found that the Russian minority is more transnationally active than the majority Estonian population.

2.4 Transnationalism and transnational people

The term "transnational" was coined by Bourne (1916) by explaining how America, as a melting pot, was becoming a *trans*-nationality instead of a nationality. The idea of transnational has developed since then and can be considered more broadly as something "extending or going beyond national boundaries" (Merriam-Webster Inc., n.d.-b). In geography transnationalism is linked to spatial movement across borders as well as the communities and networks that are formed in these environments (Gober, 2006).

Transnationalism refers to people that are tied to a multitude of places through social relations and practices, or identity (Järv et al., 2021; Vertovec, 2001). This can be for example migrants (Levitt & Jaworsky, 2007), but more recently transnational people have also been set to include a broader range of people (Glick Schiller & Salazar, 2013) like transnational working commuters (Verwiebe et al., 2017) and tourists and other temporary visitors (Silm et al., 2020). Research regarding transnationalism within geography has focused especially on transnational migration and policy (Crang et al., 2004; Levitt & Schiller, 2004; Vertovec, 2009; Walton-Roberts, 2004). However, due to data scarcity, transnational mobility patterns have been difficult to study, but with advances in technology there has been a rise in new and more complex ways of studying transnationalism (Järv et al., 2021; Silm et al., 2020). By looking at the spatial mobility of transnational people we can gain new insights in the lives of transnationals, that can in turn help explain, for example segregation patterns (Mooses et al., 2020)

2.4.1 Cross-border Mobility

Cross-border mobility is referring to mobility beyond borders, sometimes referred to as international migration (van der Velde & van Naerssen, 2010). A border is generally seen as a line forming a division between two states and Guo (2017) divides political borders into classes, whereas the highest one is the first-class borders, meaning the borders between independent nations. Border in themselves also exist in human minds and we can identify ourselves in relation to borders (van Houtum, 2005). The of studying cross-border mobility has been especially conducted through quantitative measures, like (Carpentier, 2012) who studied daily activities in the Luxembourg region based on surveys. Some more qualitative approaches have also been used, for example by the Finnish-Swedish border (Paasi & Prokkola, 2008). Cross-border mobility studies can provide a wider understanding of human mobility of transnationals (Järv et al., 2021).

2.5 Geo-located Big Data

There are many definitions of *big data*, but a common way of approaching it is by highlighting three words: *Volume*, *velocity* and *variety* (Kitchin, 2014; Laney, 2001). The volume of data is a key element for what could be considered *big* data. There's no exact line that needs to be crossed for it to be considered enough volume to be big data, and Kitchin (2014) points out that the ongoing growth of the amount of data makes it hard to define. Needless to say, we are talking about vast amounts of data. Velocity is referring to the speed the data is created at. Every moment new data is created, for example as tweets on Twitter or sensors measuring climate temperature. Variety is in this case an attempt of explaining the vast differences between data. Data can contain a wide variety of things, in our case it can be tweets, but they themselves also contain a lot of information about for example spatiality and temporality.

Geo-located big data can consist of for example mobile phone data, but in this thesis, I will use geo-located data derived from social media. Geo-located social media data provides interesting insights for mobility studies, concerning for example tourism using Twitter (Liu et al., 2018) or human mobility patterns using Weibo (Ebrahimpour et al., 2020) or Flickr (Barchiesi et al., 2015). There are several different sources to use as well

as applications, in this thesis I am applying geo-located social media data from Twitter trying to determine cross-border mobility patterns.

2.5.1 Cross-border Mobility with Big Data

Using geo-located big data sources in cross border mobility studies have been done for example with mobile positioning data (Haddawy et al., 2021; Silm et al., 2020). However, geo-located social media data provides an interesting addition to cross-border mobility research. There seem to be only a few studies done on cross-border mobility using geo-located social media data from Twitter, such as Blanford et al. (2015), Hawelka et al. (2014) and Massinen (2019).

Hawelka et al. (2014) explored global mobility patterns, increasing the understanding of human mobility depending on season and establishing the idea of using twitter data as a proxy for human mobility. Blanford et al. (2015) examined cross-border mobility patterns between Kenya and neighbouring countries.

One quite recent study of cross-border mobility is by (Silm et al., 2020), where they use mobile phone data to examine temporary population mobilities between Estonia and Finland. Another recent study found that travel restrictions due to the Covid-19 pandemic caused interesting disruptions in the mobility patterns of transnational people (Järv et al., 2021). By using geo-located big data from phones, they tracked transnational mobility patterns between Estonia and Finland and found that even though most of the people first stayed in Finland, they also moved to Estonia after the restrictions were partly lifted. This might mean that the ties to Estonia stay strong, even though the country is not visited for a while.

3 Study Setting – Minority language groups in Finland

One way to examine spatial mobility of transnational people and how these vary due to their background is to focus on minority language groups (Mooses et al., 2020). A minority language is defined by the Council of Europe (Council of Europe, 1992) as a language" traditionally used within a given territory of a State by nationals of that State who form a group numerically smaller than the rest of the State's population" and is" different from the official language(s) of that State". In this empirical case study, the study area is Finland and focusing on minority language groups that are linked to neighbouring countries: Estonia, Russia, and Sweden.

This is the definition I will be using in this thesis for Estonian and Russian. Swedish is a national language in Finland, so the second part of the Council of Europe's definition does not apply to it in this case (*Finlands Grundlag 731/1999*, 1999). Even though it is a national language, it is still a minority language in most parts of Finland, except the Åland Islands, coastal regions in the southwest and in coastal Ostrobothnia.

Finnish law does not define any minority languages and thus Russian and Estonian does not have a special legal status in Finland (Kotimaisten kielten keskus, 2021). However, the data I will be using, does not differentiate an official national language from a non-official, which means that I will not differentiate between the languages for now. Due to the different status of the language minorities, I find it interesting to study them from a perspective where the official status of the language does not interfere with how the data is collected and organized.

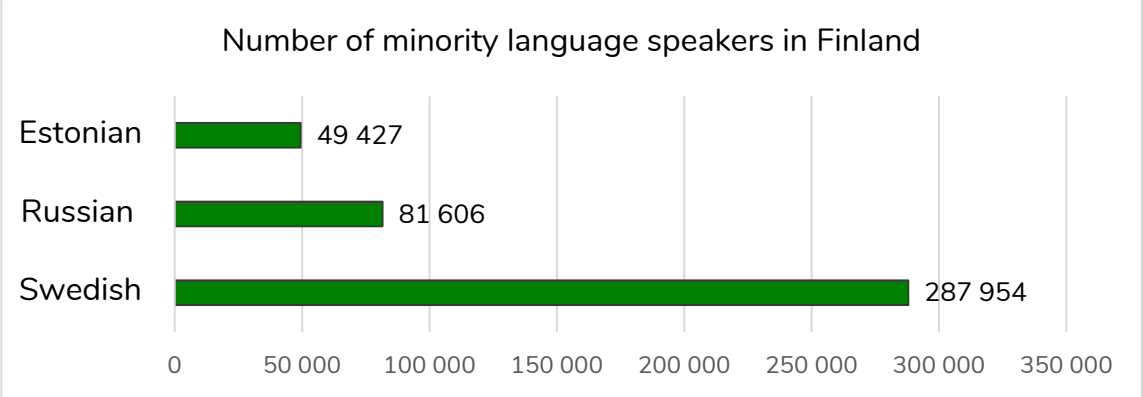

Figure 1 Minority language group sizes

From *Figure 1* we see that the number of speakers between the language groups vary. There is a total of 287 954 Swedish speakers, 81 606 Russian speakers and 49 427 Estonian speakers in Finland (Official Statistics of Finland (OSF), 2021). The number of Swedish speakers is double the amount of Estonian and Russian speakers combined. Therefore, a higher sample of Swedish tweets are to be expected in the data.

The three language minorities differ quite a lot with each other. The earliest Swedish speakers in what we now call Finland is from the 12th century and today's Swedish speakers consider themselves as part of the Finnish people (Daher et al., 2016). Swedish speakers have a strong position in society with legally guaranteed access to public services like healthcare and education (Daher et al., 2016; *Finlands Grundlag 731/1999*, 1999). There is even a completely Swedish speaking university, Åbo Akademi. Daher et al., (2016) also mentions the influence and importance of many social actors, like the Swedish People's Party which is an old political party. Some large cultural impacts by a Swedish speaker would be for example Tove Jansson and the moomins, that are seen as a national symbol for all finns (Ipatti, 2019). Stereotypically Swedish speakers are sometimes considered 'better people' by the majority population, but in reality the sociologically Swedish speakers are a mirror of Finnish speakers (Daher et al., 2016).

The roots of the Russian population stems from different historical periods, and one era in particula was the time of Russian rule between 1809 and 1917, when the language of the upper class often was Russian (Daher et al., 2016). After Finnish independece in the 1920s plenty of refugees from Soviet Russia arrived in Finland. Ingrians, from the area between St Petersburg and Estonia have arrived in Finland at multiple occasions, due to

their historical ties to Finland, especially during the second world war and in the 1990s (Daher et al., 2016; Prindiville & Hjelm, 2018). Finland only recognizes a national minority status for Russians who have lived in Finland for multiple generations, and throughout Finnish independece Russian speakers have had a struggles being accepted in society (Daher et al., 2016; Shensin, 2008) and Russian speakers have a three times higher risk of social exclusion as immigrants in Finland, compared to Estonians (Mannila & Reuter, 2009).

Estonian speakers in Finland increased in the 1990s, when Ingrians from Estonia moved, and in after the year 2004 when Estonia joined the European Union (Kährik & Tammaru, 2019). Even though the Estonian and Finnish languages are closely related, and Estonian speakers are more integrated in Finnish society than Russian speakers from Estonia, Estonian speakers have stronger ties to Estonia and are more likely to return there (Anniste & Tammaru, 2014). Estonian speakers in Finland often work in Finland but spend their leisure time back in Estonia, keeping up strong ties with Estonian culture and society.

3.1 Geographical Division

As we can see from *Figure 2* there are also clear geographical dispersion amongst the three language groups (Official Statistics Finland, 2020). The highest population share of Estonian speakers are found in the south, close by the coast. Russian speakers are more present along the eastern border of Finland and the Swedish speaking population is mainly present on the south-and Ostrobothnia coast as well as in the southwestern archipelago.

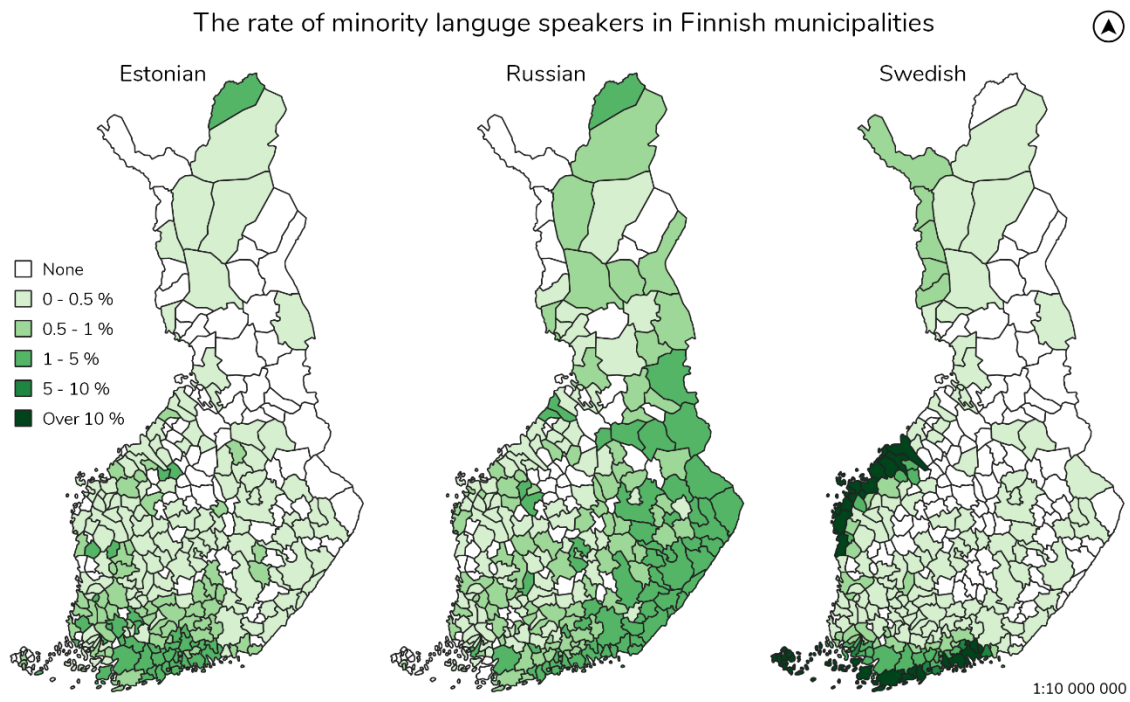

Figure 2 Language groups spatial distribution in Finland

Table 1 Municipalities with largest language minorities

Municipalities with largest language minorities								
Swedish			Russian			Estonian		
	Municipality	Speakers		Municipality	Speakers		Municipality	Speakers
1	Helsinki	36 665	1	Helsinki	18 869	1	Helsinki	10 620
2	Espoo	20 033	2	Vantaa	8 673	2	Vantaa	8 354
3	Raseborg	17 778	3	Espoo	7 409	3	Espoo	5 861
4	Vaasa	15 668	4	Lappeenranta	3 146	4	Turku	1 571
5	Porvoo	14 572	5	Turku	3 134	5	Kerava	1 144
6	Korsholm	13 326	6	Tampere	3 060	6	Kirkkonummi	1 071
7	Jakobstad	10 791	7	Lahti	2 952	7	Tampere	991
8	Turku	10 618	8	Kotka	2 499	8	Tuusula	918
9	Pedersöre	9 839	9	Joensuu	1 986	9	Nurmijärvi	909
10	Mariehamn	9 724	10	Jyväskylä	1 704	10	Lahti	884

We can see from *Table 1* that the absolute number of persons from these three groups differ(Official Statistics Finland, 2020). Helsinki has the biggest total population, so all three language groups have their biggest population there. We can clearly see that Swedish

speakers live in coastal cities, Russian speakers in Finnish cities, closer to the Russian border and Estonian speakers tend live in cities and municipalities in the south.

4 Data and Methods

4.1 Data

The data consists of 22 205 320 tweets collected by the Digital Geography Lab at the University of Helsinki with a tool called tweetsearcher (Väisänen et al., 2021). The data has been accessed through the Twitter API (Twitter, 2021).There is a total of 92 542 twitter users and out of these 43 219 have been recognised to live in Finland by using another tool also developed by the Digital Geography Lab.

Table 2 Data overview

Stage	Twitter users	Geotagged tweets
Beginning	92 542	22 205 320
Extract users living in Finland	43 129	3 980 311
Language recognized	43 126	3 897 898
Language recognized with a probability of at least 0.7	40 766	3 136 856
Tweets where the language was recognized to be Estonian, Russian, Swedish, or Finnish	35 196	2 219 105

Out of the original 22 205 320 geotagged tweets around 18 % have been tweeted by users that have been identified to live in Finland. After recognizing the languages of the tweets only a handful of users were dropped. For information about the language recognition see chapter *4.2.1*

When twitter users use geotags, they might not always be accurately placed. Some geotags might be quite coarsely placed and might refer to areas on different spatial scales. This is worth considering, especially when visually inspecting the data and when calculating distances.

The information inside the data is private and no single tweets will be published in this thesis, so that all users remain anonymous. No metrics or graphs is published where a single user can be identified. Handling of the data is in accordance with current GDPR regulations.

4.2 Methods

In this section the main methods used are explained. *Figure 3* presents an overview of the workflow. My first and foremost tool in this thesis is Python. To handle big data, I've been using PostgreSQL for handling large quantities of data. The scripts I've used are available on GitHub and they are in many cases modifications and adaptions of already made scripts by the BORDERSPACE-project and the Digital Geography Lab of the University of Helsinki (Aagesen, 2021; Massinen, 2019).

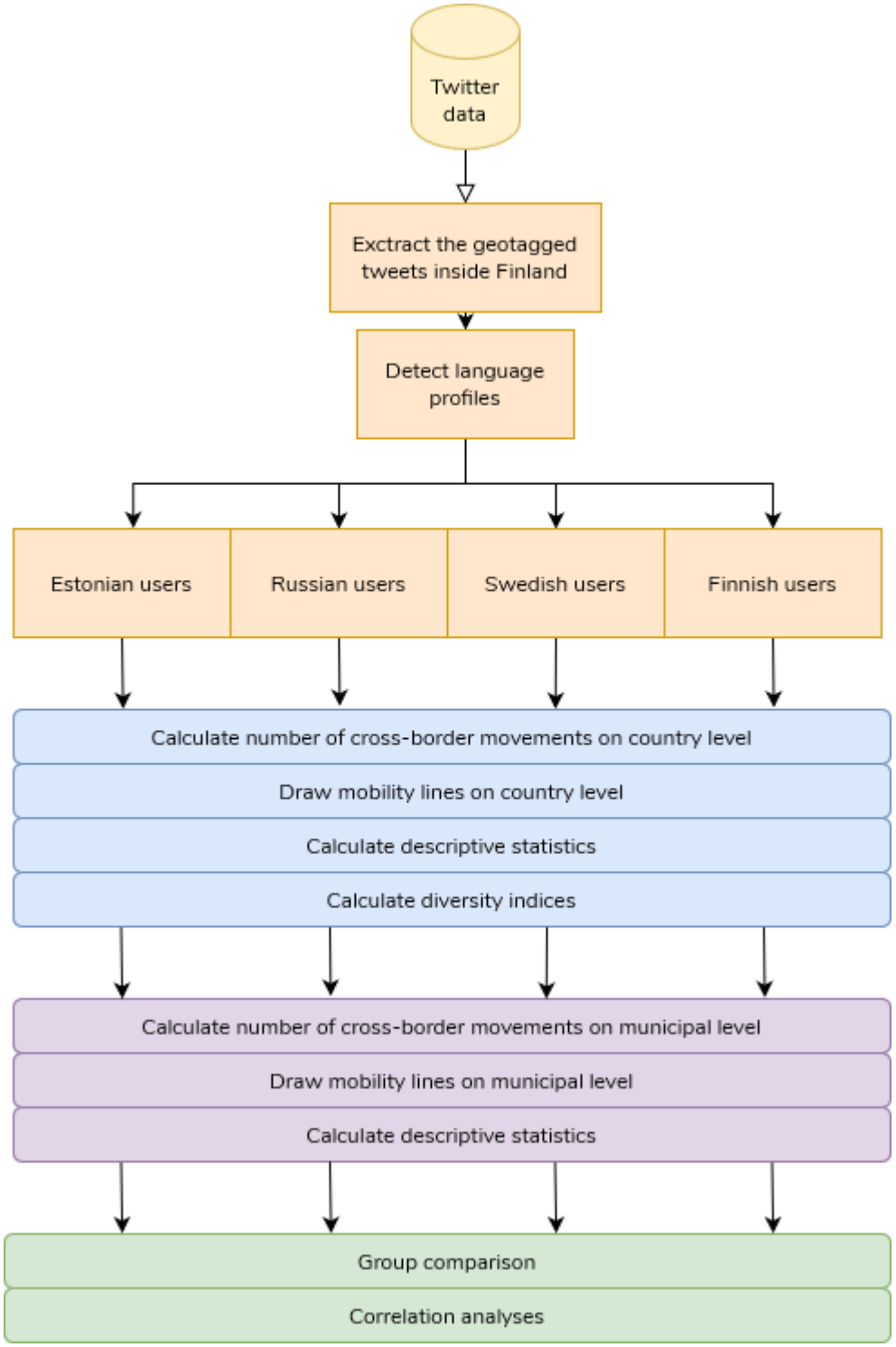

Figure 3 Flowchart over methods

4.2.1 Language Identification Methods

Tweets contain short pieces of texts and with these texts it is possible to draw conclusions about the language used in the tweet. By automating the process, an immense amount of time is saved. A common tool for recognizing languages is the Python library fastText (Heikinheimo et al., 2018; Hiippala et al., 2020). I am using a modified script that is develop by the Digital Geograhpy Lab at the University of Helsinki (Digital Geography Lab, 2020). By using a pre-made model I identify the languages of tweets and extract the tweets that are interesting for this topic: Estonian, Russian, Swedish, and Finnish.

A threshold of a 70 % language probability is picked. All the tweets with a lower certainty are disregarded. The same threshold has been used in similar studies where language recognition of tweets has been made (Hiippala et al., 2020).

After assigning each tweet a language probability for each language, the tweets are grouped by each user. For each user an average language probability for each language is calculated. Then the users are assigned one language group that they represent.

As a threshold value 10 % is chosen. This means that if a user's tweets are at least 10 % in Estonian, that user is recognized as an Estonian-speaking user. For the same user not to be included in multiple language groups, a prioritization was done in the following order: Estonian, Russian, Swedish, and Finnish. If a user's tweets consist of 11 % Estonian tweets and 11 % Russian tweets, that user is only included in the Estonian language group. English is excluded from this language group matching, since it is used so widely amongst Twitter users and works as a lingua franca and is thus hard to match to a certain group of people.

Of a total of 3 980 311 tweets made by people residing in Finland a language is recognized with a certainty of 70 % for 3 136 856 tweets. Out of these tweets, 2 219 105 is in either Estonian, Russian, Swedish, or Finnish by 35 196 different users. With the 10 % threshold the total number of users is 31 042. The idea behind this line of thinking is that if a user's uses a minority language for at least 10 % of their tweets, that user probably knows that language. Then it is assumed that that user either is a native speaker of that language or that they might have ties to that language group in other ways. This might not be the case

for every particular user, but the assumption is made that this is the case for the most part. The division between language groups can be seen in *Table 3*.

Table 3 Comparison between share of Twitter users and population shares

	Estonian	Russian	Swedish	Finnish
Number of users	225	653	1 609	28 552
Share of users	0.6	1.9	4.6	81.2
Share of total population*	0.9	1.5	5.2	86.9
*(Official Statistics Finland (OSF), 2020)				

The population of Finland consists of 0.9 % Estonian speakers, 1.5 % Russian speakers, 5.2 % Swedish speakers and 86.9 % Finnish speakers (Official Statistics Finland (OSF), 2020). These numbers are used as a guideline for choosing the threshold of how likely it is that a user would be part of a language group. There are of course some caveats with this data, and it is not perfectly matched with the actual population. Some users might also identify with multiple language groups, but in this case the smallest minority is prioritized.

4.2.2 Cross-Border Mobility Identification

To identify cross-border mobilities for a user the geotags were extracted from each tweet. For each user the tweets are sorted chronologically and then a line is drawn between the first tweet to the second tweet, from the second to the third and so on. If a line was not originated nor destinated to Finland, the line was dropped from the data. This whole process is made for each language group separately.

Identifying cross-border mobility is also applied on a municipal level and then only tweets originated and destinated in Finland are included. For the municipal level cross-border mobility identification the distance was also calculated. The distance is Euclidean and locational accuracy is worth considering.

4.2.3 Calculating Statistics and Indices

To try to understand the mobility diversity of users I calculate the number of unique countries visited per user. To get a more nuanced picture of the mobility diversity, I've used some indices familiar from ecology. The richness of diversity could be explained by

the number of visited countries for the minority language group, but it doesn't say anything about for example evenness and dominance. For this, I turn to indices.

The Shannon entropy has been used to explain species diversity in an area (Shannon, 1949; Spellerberg & Fedor, 2003), but for this thesis I am using the index to compare the diversity of mobilities between language groups. I use the Shannon entropy to see how evenly the users of one language group tend to travel to different locations. With this I try to see, if it is a common tendency for users within one language group to have more diverse travel patterns or if the travels are focused to one country. For each user the Shannon entropy is calculated and then a mean of the whole language group is presented.

For calculating of the Shannon entropy, I use the python library ecopy and the formula is demonstrated in *Figure 4* (Lemoine, 2015).

$$H = -\sum_{1}^{k} p_k \log p_k$$

Figure 4 Shannon entropy formula

where *k* in this case is the proportion of visits to a particular country.

Another popular way of measuring diversity is Simpson's index (Simpson, 1949). It takes into account both richness and evenness. There are a few different versions of the Simpson index and in this study, I am using the one referred to as Simpson's Reciprocal Index. The lowest possible index value is 1 and the highest possible value is the highest number of unique countries. This index can give us an idea of the travel diversity of the language groups. I use the python library ecopy to count this index which and the formula for the calculation is visible in *Figure 5* (Lemoine, 2015).

$$D = 1 - \sum_{1}^{k} p_k^2$$

Figure 5 Simpson's Reciprocal Index formula

Where *k* is the proportion of travels to a unique country.

4.3 Analysis

4.3.1 Linkage to Neighbouring Country & Cross-border mobility

To answer the first research question concerning language minorities spatial connectedness with neighbouring countries I turn to some user level calculations. For each user of a language group, I calculate the share of travels to their respective neighbouring country with a corresponding majority language from the total number of cross-border tweets. For each user I also calculate the share of travels to their corresponding majority language country. An average is calculated based on the shares per user. This number will give an indication of the concentration of travels to the neighbouring country.

Another way of exploring the spatial relationship between neighbouring countries is by drawing lines between each user tweets in a chronological order, to see how the user has moved around (Aagesen, 2021; Massinen, 2019). I am using a method provided by Massinen (2019). By examining the movement lines, we get a diverse picture of the places people are traveling between. These movement lines also indicate travels to other countries, which give us an idea of other the concentration of movements to a single country in relation to other countries.

4.3.2 Domestic Mobility

The research question about the domestic movements of the language groups is also being addressed in a similar fashion. For each user I calculate the number of unique municipalities they've visited within Finland. This is to explain how mobile the language groups are domestically. However, municipality areas vary, and some might still be very mobile, but they just always go to the same municipalities. That's why I am also calculating the Euclidean mean distance per user for each language group. This distance indicates the mobility, regardless of municipal boundaries. To get a better understanding of the hotspot's users move in between, I also visually inspect the movement lines within Finland.

4.3.3 Language Group Comparison

In my research questions I also mention how I am interested int the diversity of the mobility patterns between groups. In my efforts to explain the diversities of movements between language groups I use indices familiar from Ecology. Both Shannon's entropy and Simpson's index of diversity are used to explain the diversity of mobility patterns.

One central piece of this thesis is to find out if there are difference between the mobility patterns of the language groups. To determine significant differences between groups we turn to a one-way ANOVA test in SPSS (IBM Corp., 2019). For each of these variables the one-way ANOVA is calculated: number of unique countries, number of unique municipalities, Shannon entropy, Simpson index, domestic distance mean, and the share of travels to majority language country. For the one-way ANOVA to be considered the homogeneity of the variances must not be significant. If they are significant, the Welch-correction need to be considered. It tests the equality of means and if it is significant, we continue with a proper post-hoc test. The one-way ANOVA explains overall differences, it does not tell us between which groups the differences are. For this purpose, we use a post hoc test and in this case one called Games-Howell, which can be used when homogeneity of variances is violated. The Games-Howell test indicates where the differences between groups is significant.

4.3.4 Linkage between Cross-border and domestic mobility

To understand the connection between how mobile a user is across country borders and within the country I turn to Pearson correlations. By doing a correlation analysis between the cross-border variables (number of unique countries visited, Shannon entropy and Simpson's index) and the domestic mobility variables (number of unique municipalities visited and the mean of kilometres travelled within the country), a connection between the mobility patterns across and withing borders can appear. These analyses can give some insights into the nature of mobility for the language groups, for example if a person that travels a lot to other countries also is travelling a lot within the country.

5 Results

The result section is divided into four parts. One chapter is focusing on the cross-border movements abroad and the second one on movements within Finland, on municipal level. In the third part the languages groups are compared with a one-way ANOVA, to find significant differences. Lastly, I compare the transnational variables with the domestic ones to see if there is any correlation between a high mobility across Finnish borders and within Finland. Keep in mind that in all the cross-border maps, all movement lines that do not have their origin or destination in Finland are excluded.

5.1 Cross-border Movements

To illustrate the differences for longer cross-border journeys we can divide the movements between continents. In *Figure 6* the share of travels to each continent is displayed by language group. In this figure, Russia is considered Europe. Naturally the biggest share of travels is in Europe for all language groups. However, for Russian speakers almost 93 % of cross-border travels are in Europe, while the corresponding number for Finnish speakers is only around 80 %. Instead, Finnish speakers seem to travel more often to North America and Asia. For Swedish speakers 85 % of cross-border travels happen in Europe, and after that mainly in North America and Asia, just like Finnish speakers. Estonian speakers' cross-border travels are more concentrated to Europe with almost 91 % of travels situated there. After Europe the most popular continents for Estonian speakers are Asia and North America. As a sidenote, we can see that Russian have a slightly lower share of travels to North America than the other language groups.

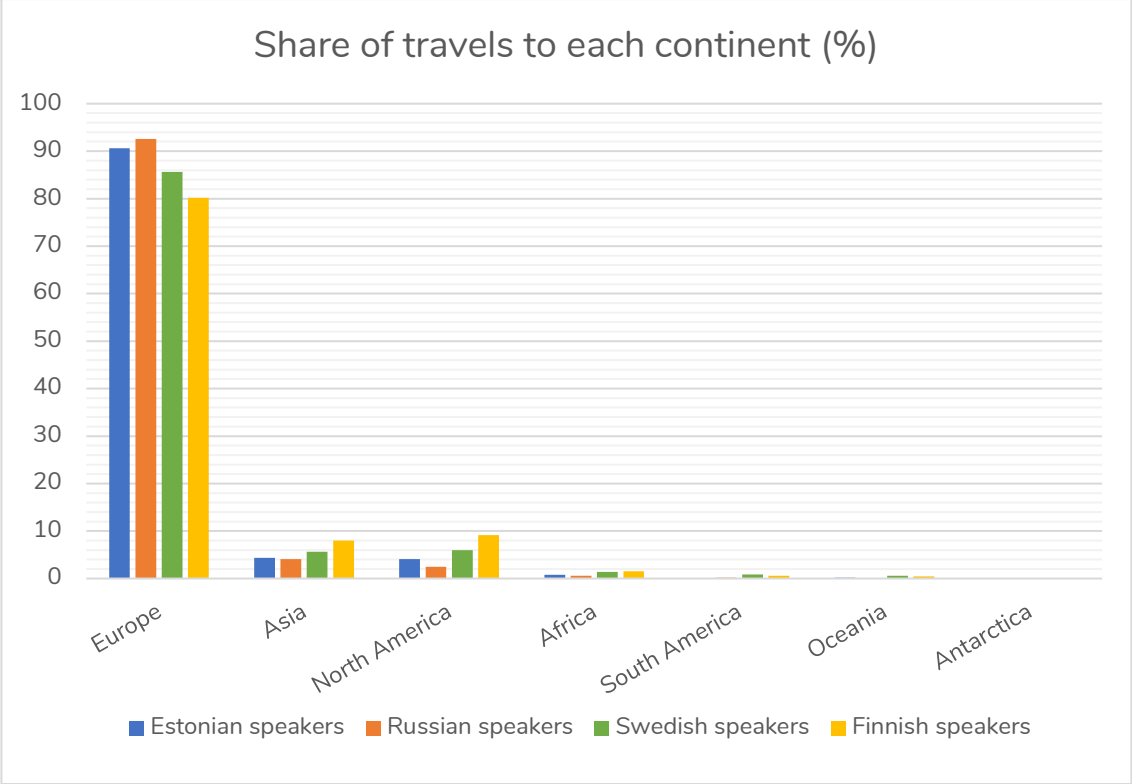

Figure 6 Bar chart over share of visits to different continents for each language group

The cross-border movements between countries for each language group is presented in *Table 4*. We can see that the amount of data for each language group differs quite a lot, as a reflection of the demographics of Finland. For comparison of each language group, the number of cross-border movements per user is presented. The Estonian and Finnish language groups have quite similar amounts of cross-border movements, Swedish speakers a bit more and Russian speakers have the highest number. The absolute number are shown to shed light on the differences in sample size between the groups.

To understand the distribution of share of trips to countries with a majority language we look at *Figure 7*. Here we see that the three language minorities share to trips to majority language countries differ. Estonian speakers visit Estonia often and for the middlemost Estonian speaker half of the cross-border movements are between Estonia and Finland. The variation for Russian speakers is wide, compared to the other minorities and the mean and median are close to each other, around 0.26. This means that some Russian speakers visit Russia often while other almost never visit Russia. For Swedish speakers the travels between Sweden and Finland are on average modest compared to the other two languages.

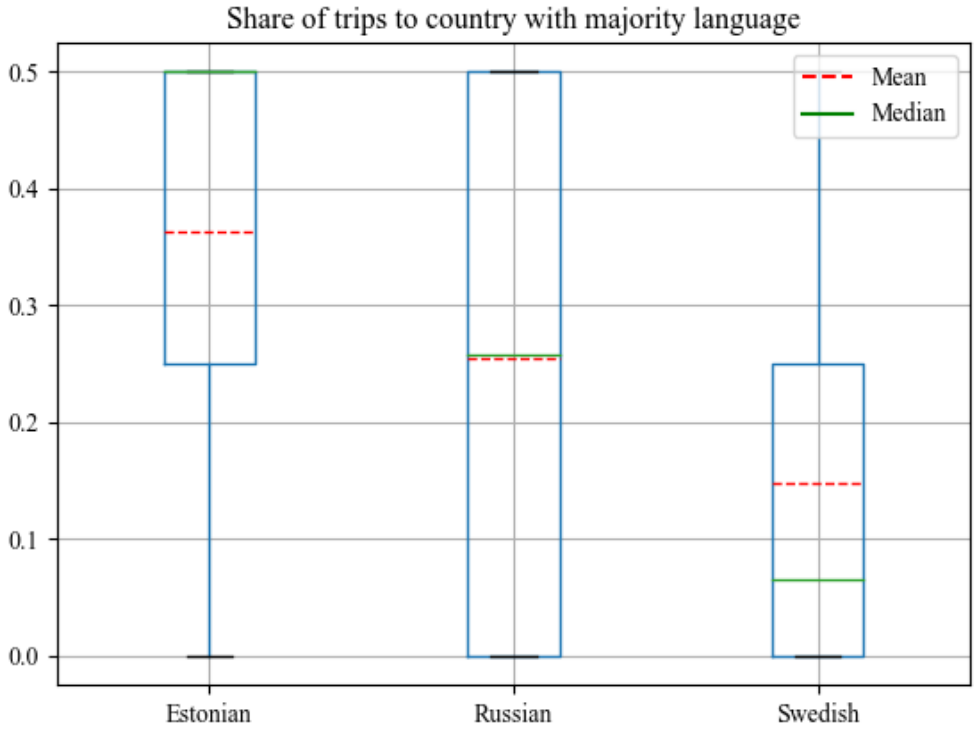

Figure 7 Boxplot over the share of travels to neighbouring majority language country

As an example of the distribution of tweets *Figure 8* is showing what the movement lines of Russian speakers residing in Finland look like on a global scale. With this view, it is possible to see that also places outside of Europe are visited, east and south Asia as well as well as parts of North America. The main concentration is naturally between Finland and Russia and Finland and the rest of Europe. Similar maps for Estonian speakers, Swedish speakers and Finnish speakers can be found in *Appendix A*.

Table 4 Cross-border movement data overview

	Estonian	Russian	Swedish	Finnish
Users	225	653	1 609	28 555
Users with cross-border movements between Finland and another country	140	384	912	12 423
Cross-border movements	1 131	4 606	9 782	106 693
Cross-border movements per user	8.08	12.00	10.73	8.59

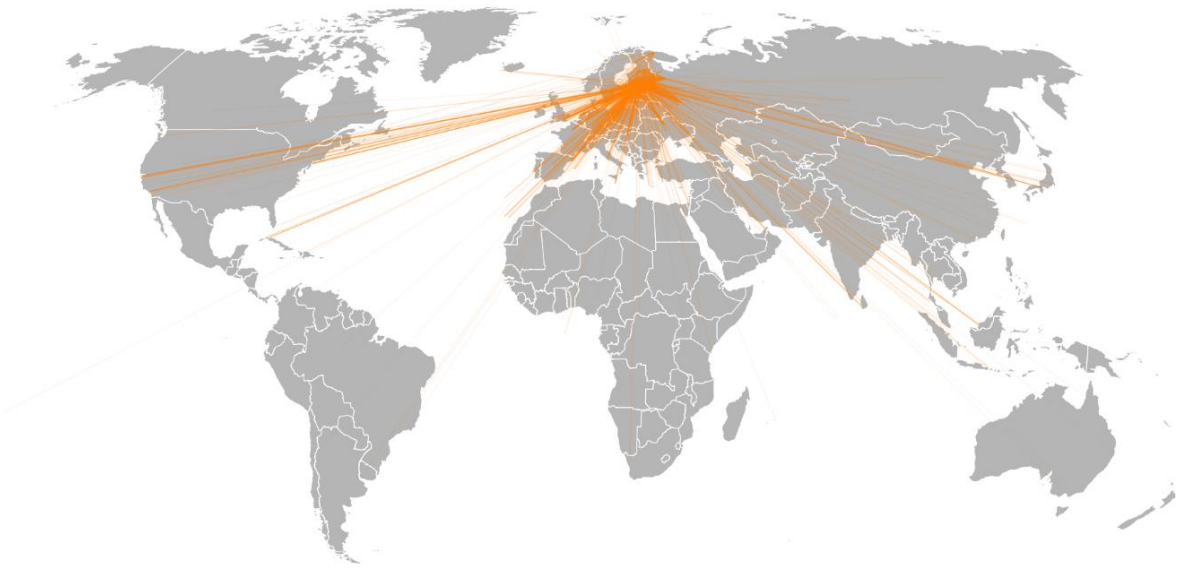

Figure 8 Global movement of Russian speakers residing in Finland

5.1.1 Linkage to Neighbouring Countries

The relationship between the language group and their respective neighbouring countries with their language as a majority language, is displayed in *Table 5*. Bear in mind that the total amount of cross-border movements between language groups varies: Estonian speakers with 1 131 cross-border movements, Russian speakers with 4 606 and Swedish speakers with 9 782. To be able to compare the three language groups we look at the cross-border movements to the neighbouring country with the corresponding language majority per person. According to this data the Estonian and Russian speakers have very similar values, a little over five visits per person. Swedish speakers have a somewhat lower number with only 3.16 visits to Sweden per person. The difference in the concentration of cross-border movements to their respective neighbouring country on the other hand is evident. Out of all the cross-border movements by Estonian speakers, 68 % are to Estonia. For Russian speakers and Russia, the number is 45 % and for Swedish speakers and Sweden it's 29 %. Looking at the percentage of trips to the majority language country *per person* we see that on average an Estonian speaker's trips go to or from Estonia 36 % of the time. The figure for Russian speakers to and from Russia is 25 % and for Swedish speakers to Sweden a bit less than 15 %.

Table 5 Language group comparison of cross-border travels and share of movements to majority language country

	Estonian	**Russian**	**Swedish**	**Finnish**
Cross-border movements to majority language country	769	2079	2880	NA
Cross-border movements to majority language country *per person*	5.49	5.41	3.16	NA
Proportion of cross-border movements to majority language country out of *all* cross-border travels	68.0 %	45.1 %	29.4 %	NA
Proportion of cross-border movements to majority language country *per person*	36.3 %	25.4 %	14.8 %	NA
Unique countries visited	44	83	121	194

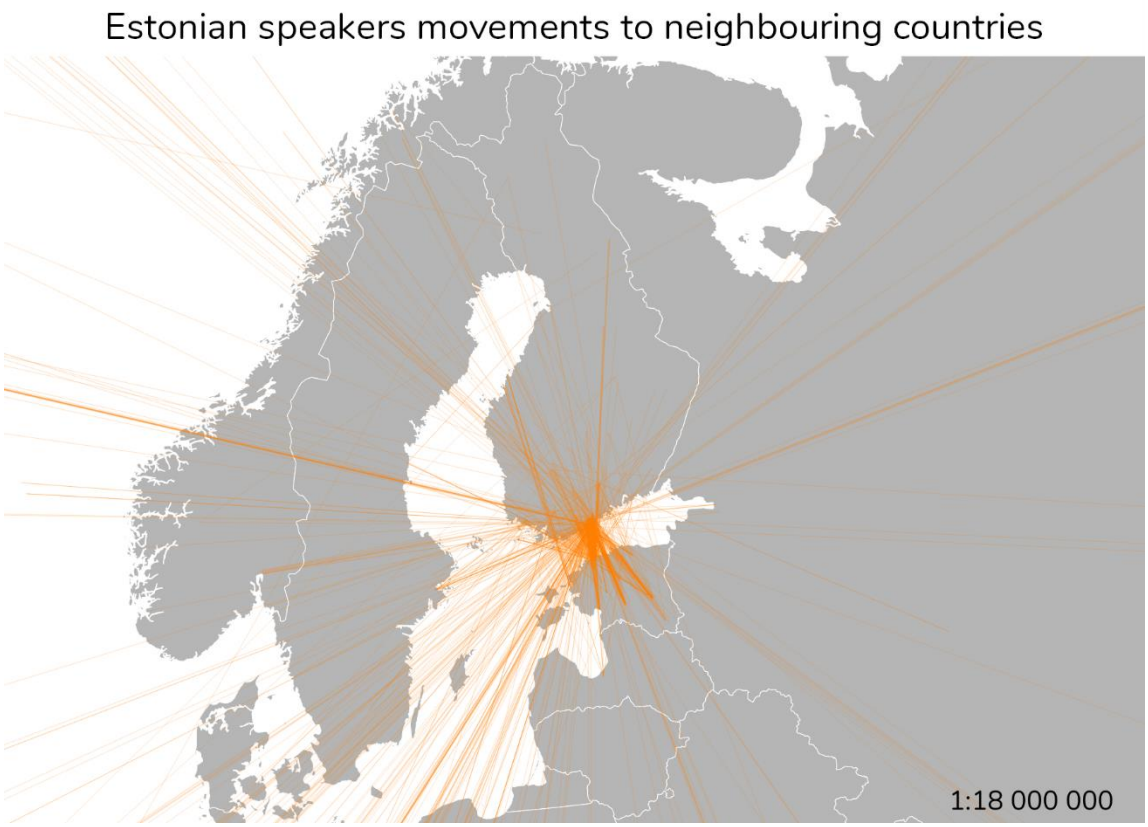

Figure 9 Mobility of Estonian speakers residing in Finland between Finland and its neighbouring countries

As an overview of the Estonian speakers' movements to neighbouring countries in Northern Europe we look at *Figure 9* where we can see a clear concentration of movement lines across the Finnish Bay. The movement lines show that Estonian speakers seem to be traveling mostly between Estonia and southern Finland, only a few lines show activity in more Northern regions.

A more detailed map of Estonia and southern Finland show where the most important mobility patterns occur in *Figure 10*. In Estonia, Tallinn is the most visited city. Also, other major cities appear on the maps. Tartu, Viljandi and Pärnu are the most prominent ones, besides Tallinn. On the Finnish side the Helsinki region is the main destination. There are also some movement lines going to for example Lahti, outside the metropolitan area. We can see that the major cities in Estonia, with some exceptions, (like Narva) are showing up on the map, while on the Finnish side the dominant area would is clearly Helsinki.

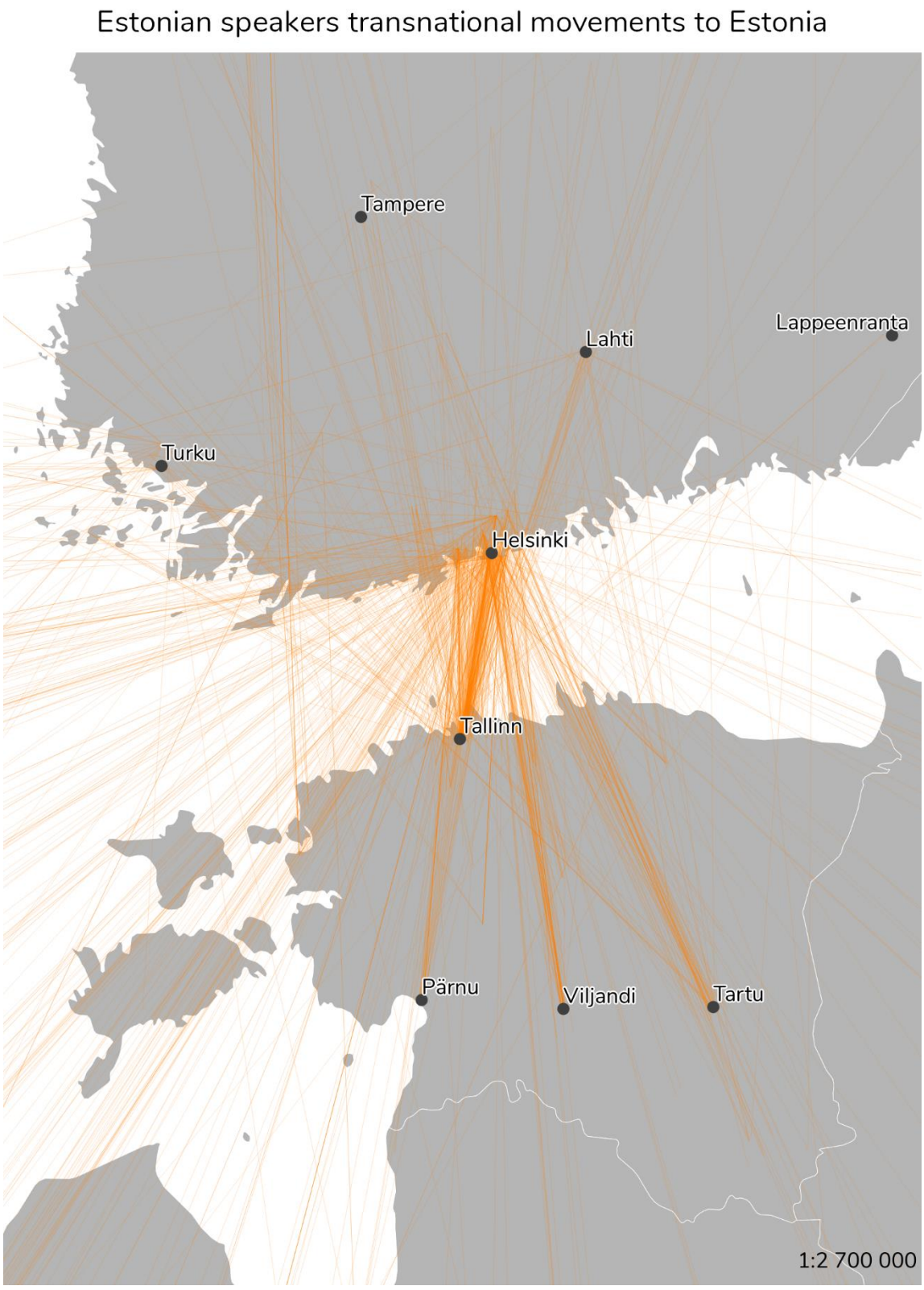

Figure 10 Estonian speakers' movements between Estonia and southern Finland

The movement lines of the Russian language groups can be seen in *Figure 11*. Due to the lengthiness of Russia only the most visited part is shown. The highest concentration of movement lines is also in the more densely populated south of Finland. Russian speakers seem to travel between Helsinki and St Petersburg. As well as St Petersburg and other minor cities in the eastern part of Finland, like Lappeenranta and Imatra. From Helsinki we can also see some quite strong movement lines going to the capital Moscow. The biggest cities are for obvious reasons clearly visible, but we can also see some movements to smaller places in Russia, like Murmansk by the Arctic Sea and Northern Finland. Some lines also indicate movements between Petrozavodsk, the capital in the Karelian Republic and southern Finland. Russian speakers living in Finland also seem to visit Stockholm in Sweden. Too see more of where the movement lines go within northern Europe please see *Appendix B.*

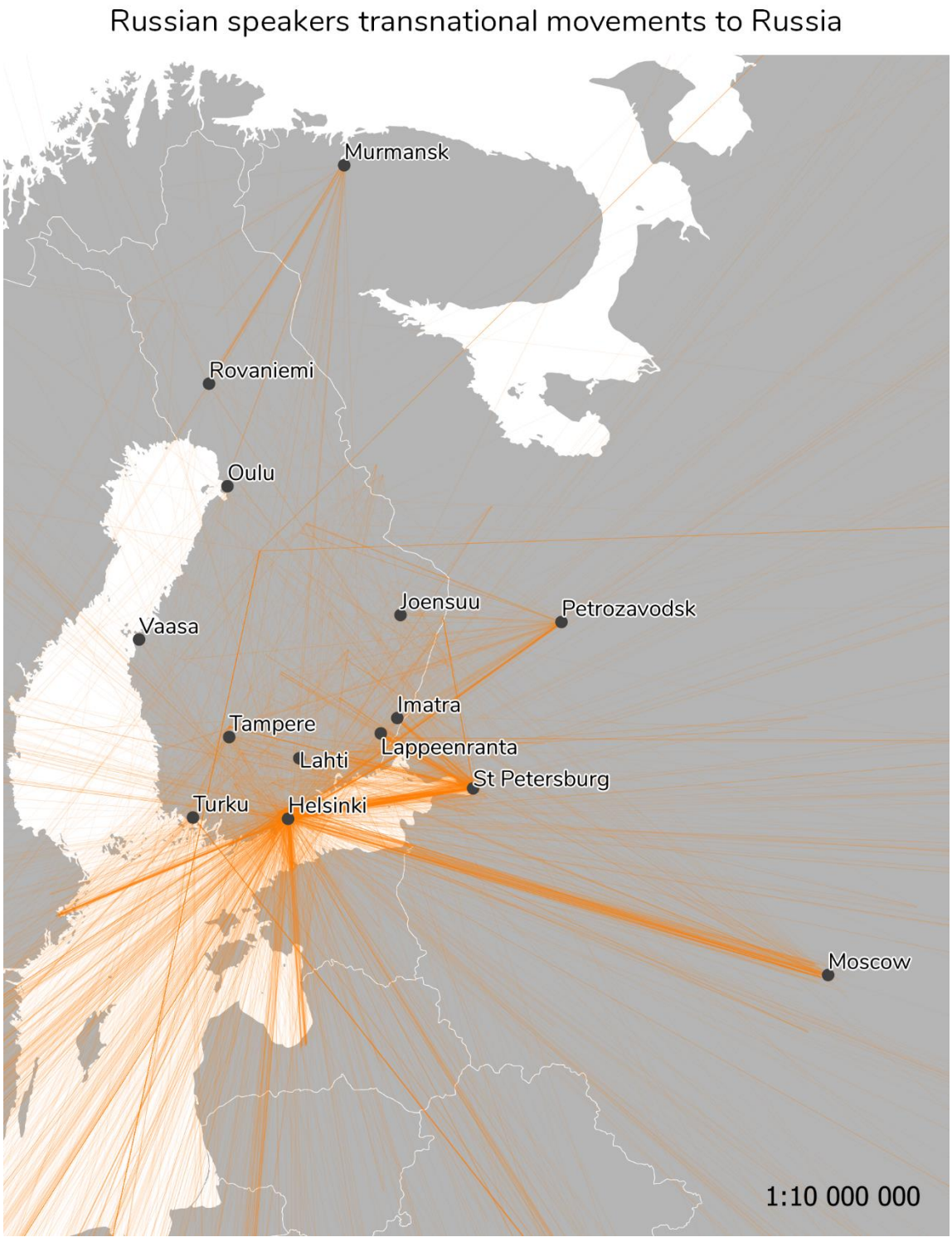

Figure 11 Mobility of Russian speakers residing in Finland between Finland and its neighbouring countries

The cross-border movements for Swedish speakers can be seen in *Figure 12*. In comparison to the previous maps, we can see that there are more users leading to more movement lines. Comparing to the previous map, we now see that the focus has shifted westwards and that it is not only the southern part of Finland being hotspots but also the coast of Ostrobothnia, specifically around Vaasa. For Swedish speakers Stockholm is a highly visited location, both from southern Finland and from Ostrobothnia. The second largest city in Sweden, Gothenburg is also a frequently visited place. Other major Nordic cities also show up on the map, such as Oslo and Copenhagen. There's also a density of movement lines going towards other parts of western Europe. One destination particularly strong seems to be between London originating mostly from Helsinki but also from Ostrobothnia. From Vaasa we can also see movement lines going across the sea to Umeå on the Swedish side. From all Swedish speaking regions, the movement lines across the eastern borders are quite sparse.

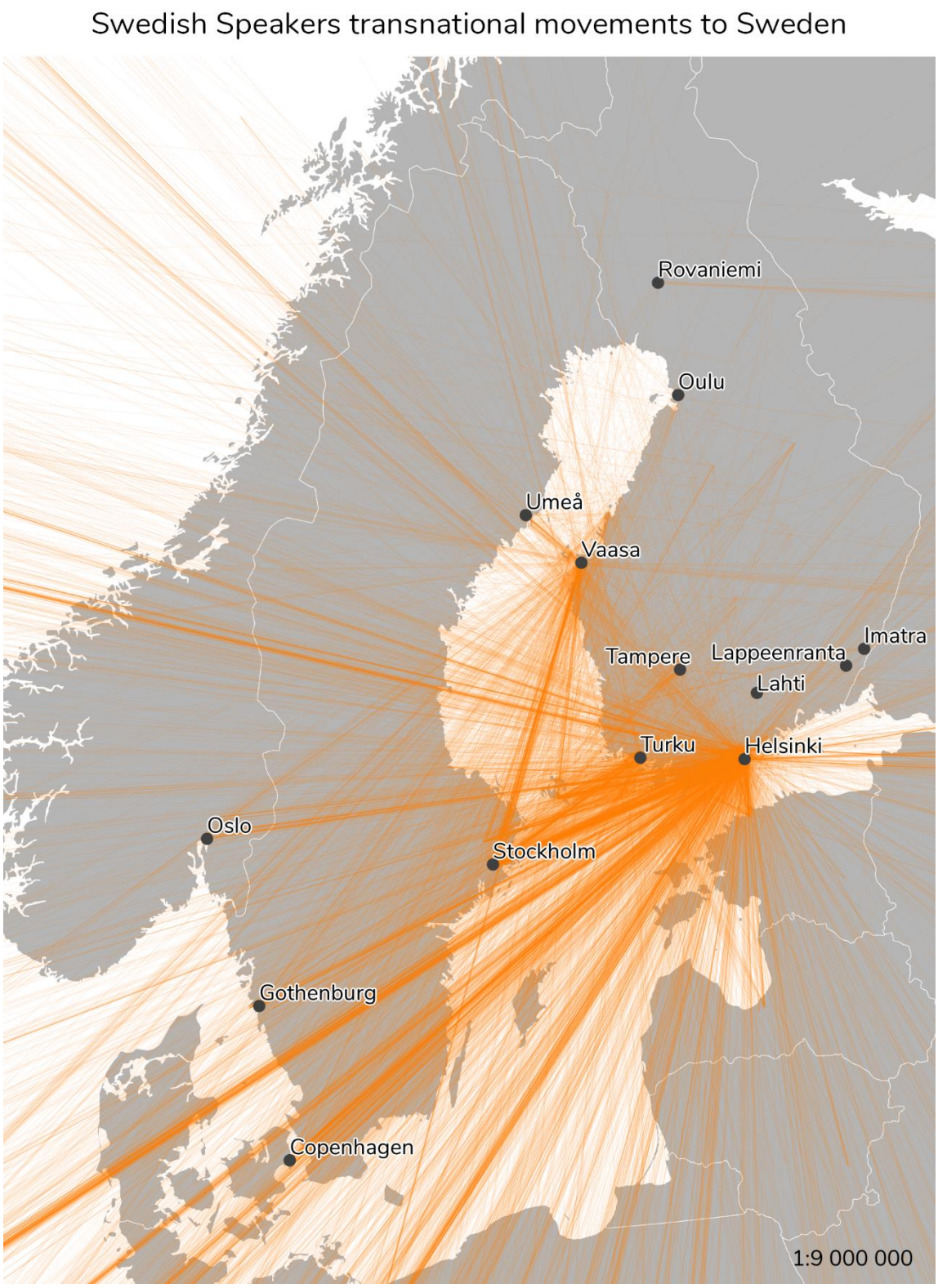

Figure 12 Mobility of Swedish speakers residing in Finland between Finland and its neighbouring countries

5.1.1.1 Cross-border Mobility Diversity

The Shannon entropy is calculated for each user in the language group separately and then a mean for each language group is derived. The means can be viewed in *Table 6*. The highest Shannon entropy mean is by the Swedish speaking minority (3.05), followed by Finnish speakers (2.94), then Russian speakers (2.62) and lastly Estonian speakers (1.70).

Table 6 Comparison of mean of diversity metrics

	Estonian	Russian	Swedish	Finnish
Mean of Shannon entropy	1.88	2.62	3.05	2.94
Mean of Simpson index	1.70	2.25	2.71	2.68

For understanding how the different language groups are placed on the Shannon entropy spectrum, please see *Figure 13*. The variation is largest amongst Swedish, then Finnish and then Russian speakers. The Estonian speakers' median and mean differ quite a lot. Because of the small sample size less outliers skew the data. The minimum Shannon entropy value is 1 and it seems like that is very common amongst Estonian speakers. This means that there are plenty of people with a low travel diversity.

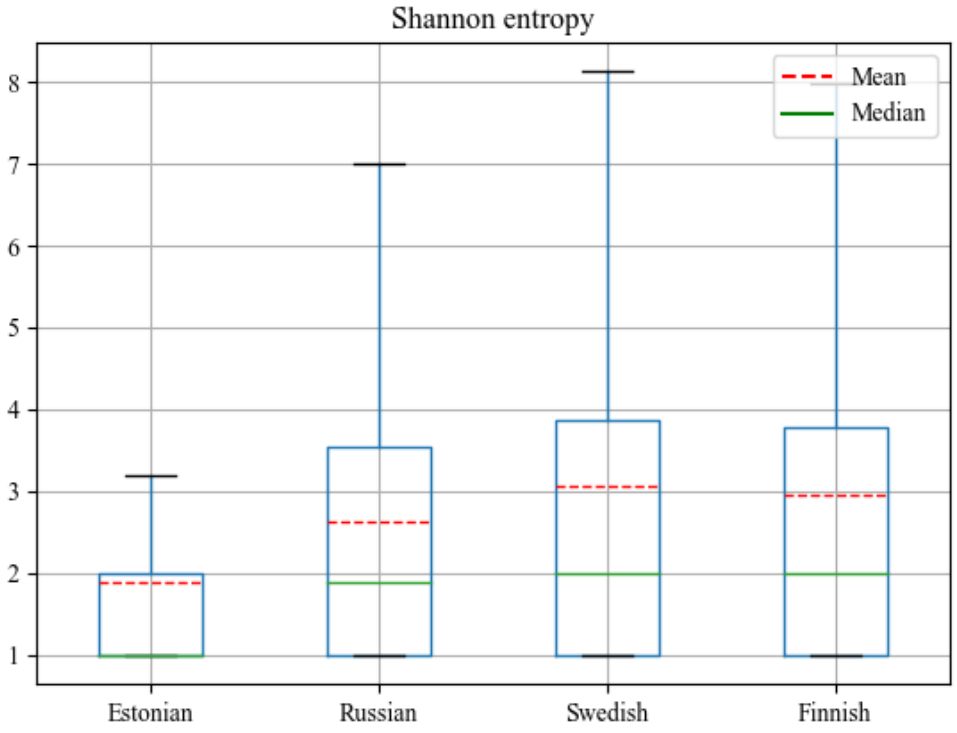

Figure 13 Boxplot over Shannon entropy values

When looking at the distribution of the Simpson's Reciprocal index in *Figure 14* we can see similar tendencies like in the Shannon entropy figure. The Simpson index value minimum is 1 and the highest possible value is the unique number of countries. There seems to be a high abundance of low diversity Estonian speakers, when considering that the mean is 1. Swedish and Finnish speakers have very similar data variations and Russian speakers are in the middle. The mean index values that are seen in *Table 6* seem to be differing from each other with similar intervals as for the Shannon entropy.

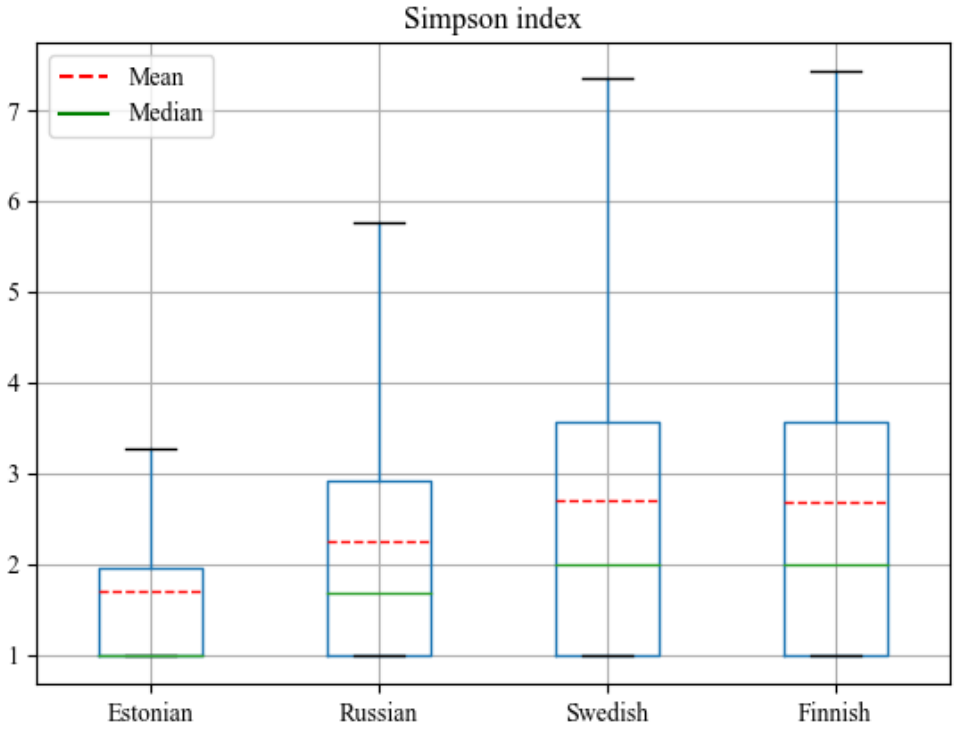

Figure 14 Boxplot over Simpson's index values

5.2 Domestic Mobility

To be able to compare the cross-border movements with domestic mobility we can look at movements between municipalities. Out of the roughly 300 municipalities in Finland they vary plenty in size. Especially in northern Finland the areas of municipalities are quite large in comparison to southern Finland. Therefore, a mean Euclidean distance is also calculated for each user since it does not abide by municipality borders.

Table 7 Municipalities visited per person

	Estonian	Russian	Swedish	Finnish
Users with cross-border movements as well as domestic movements	93	267	783	10 861
Unique municipalities visited per user	4.76	5.51	7.76	8.40

In *Table 7* we can see that the number of users that both have been traveling across-borders and within Finland are dropping meaning that the sample sizes for the smaller language groups are very tiny. The amount of domestic cross-border movements varies quite a lot. The last row in the table is worth comparing. Estonian speakers have the lowest number of unique municipalities visited per user, less than 5, while Finnish speakers have the highest number, 8.4. Swedish speakers visit on average 7.76 municipalities and for Russian speakers the number is 5.51.

From *Figure 15* we can see the variation in number of unique municipalities. We see that the median for Estonian speakers and Russian speakers is the same, while the average is a bit higher for Russian speakers. We also see bigger variation amongst Russian speakers. Swedish and Finnish speakers also share the same median while on average the number of unique municipalities visited is higher amongst Finnish speakers

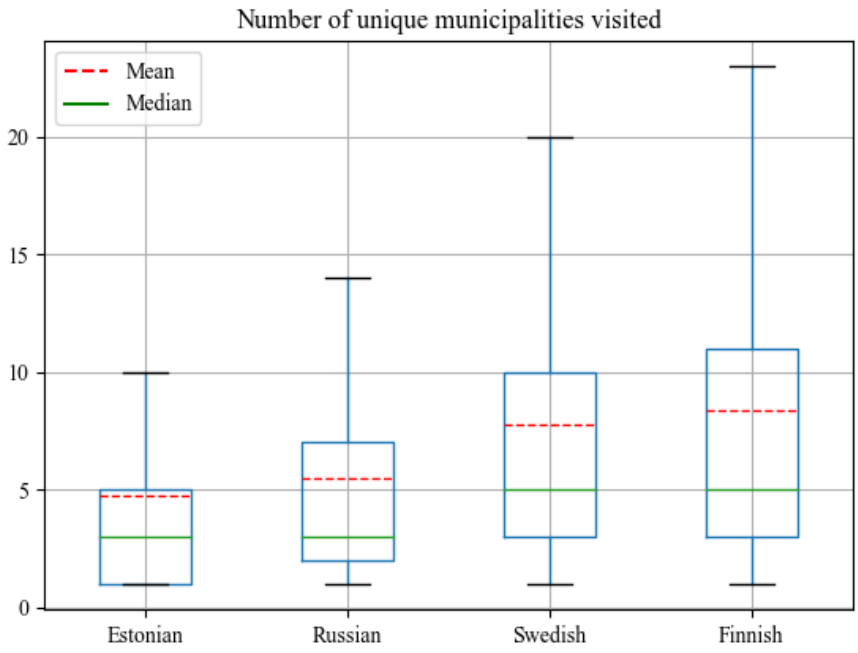

Figure 15 Boxplot over the number of unique municipalities visited

5.2.1 Mobility between Municipalities

To discover differences between the language groups we look at *Table 8*. For each user a total mean distance of all domestic travels between municipalities was calculated and then a language group mean was calculated from that. It will hereby be referred to as domestic distance mean. The distance travelled by Estonian speakers is a lot lower compared to the other language groups. Russian and Swedish speakers only differ with a few kilometers while Finnish speakers have a higher mean. This could indicate that Estonian users are more sedentary and do not travel within Finland as much as the other language groups. The location of the population is also good to keep in mind. Finnish speakers are found everywhere in Finland, while the other three groups are more concentrated, perhaps leading to different travel habits with shorter distances.

In *Figure 16* we can see the variations in domestic distance mean between the language groups. Russian speakers seem to have the largest variation, with some people travelling a lot within Finland and some very little. We can see that the upper whiskers are at the same height as for Finnish speakers even though the averages vary. The Swedish speaking population is very concentrated to coastal areas and perhaps this might explain why they are travelling shorter distances than Finnish speakers. Estonian speakers are not as dispersedly placed in box plot as the rest of the language groups. This could indicate that the average Estonian speaker do not travel as much within Finland as the other language groups.

Table 8 Domestic distance mean for each language group

	Estonian	Russian	Swedish	Finnish
Mean of overall distance travelled within Finland (Domestic distance mean)	77.3	120.5	127.5	138.4

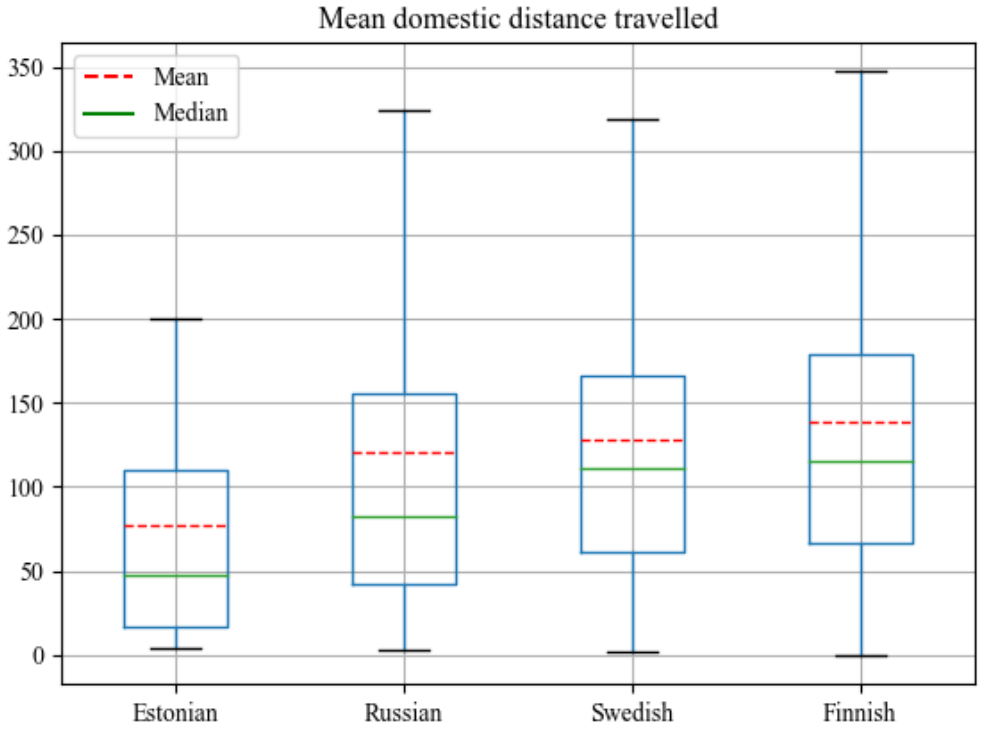

Figure 16 Box plot over domestic distance travelled

To get a spatial overview of where these domestic movements happen, we can look at *Figure 17* and *Figure 18*. Keeping in mind that the number of movement lines differ between groups, we can still somewhat understand where the three minority groups move. Estonian speakers move around in Finland mainly in the metropolitan area and from there to Turku, Tampere, and Lahti. Russian speakers also have similar pattern to start with, but there are definitely more trips being made to northern Finland, perhaps for touristic purposes. It is also visible that Russian speakers have a stronger connection to the east and people seem to travel between the metropolitan area and Lappeenranta and Imatra.

As for the Swedish speakers the movement lines are naturally more in Swedish speaking regions, as we can see from *Figure 1*. There are very strong lines on the south coast. All the way from Porvoo to Turku. In Ostrobothnia we see a concentration of lines from and to the bigger cities with Swedish speaking populations. There seems to be a lot of movement between southern Finland and Ostrobothnia as well, resulting in a very different map, compared to the two previous ones. The Finnish map a random sample of

movement lines of Finnish speakers, since bringing all lines on the map resulted in a completely unreadable map. With the random sample we can see that movements are strongest between bigger cities. Helsinki is often the origin and destination for Finnish speakers as well.

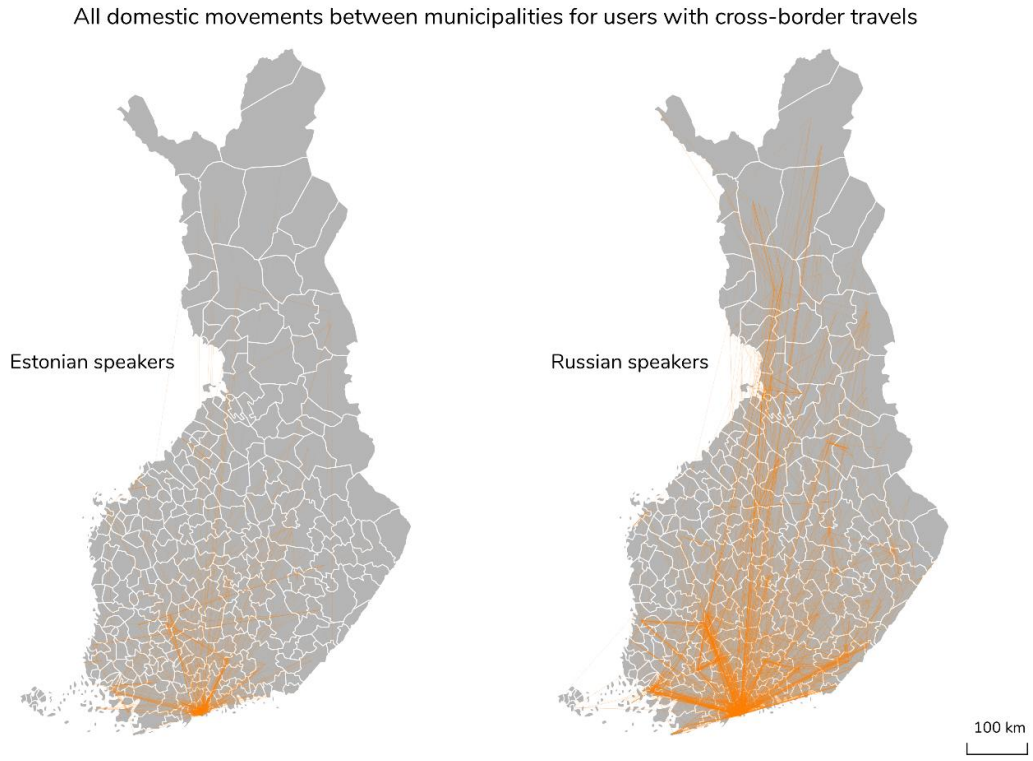

Figure 17 Domestic movements of Estonian and Russian speakers

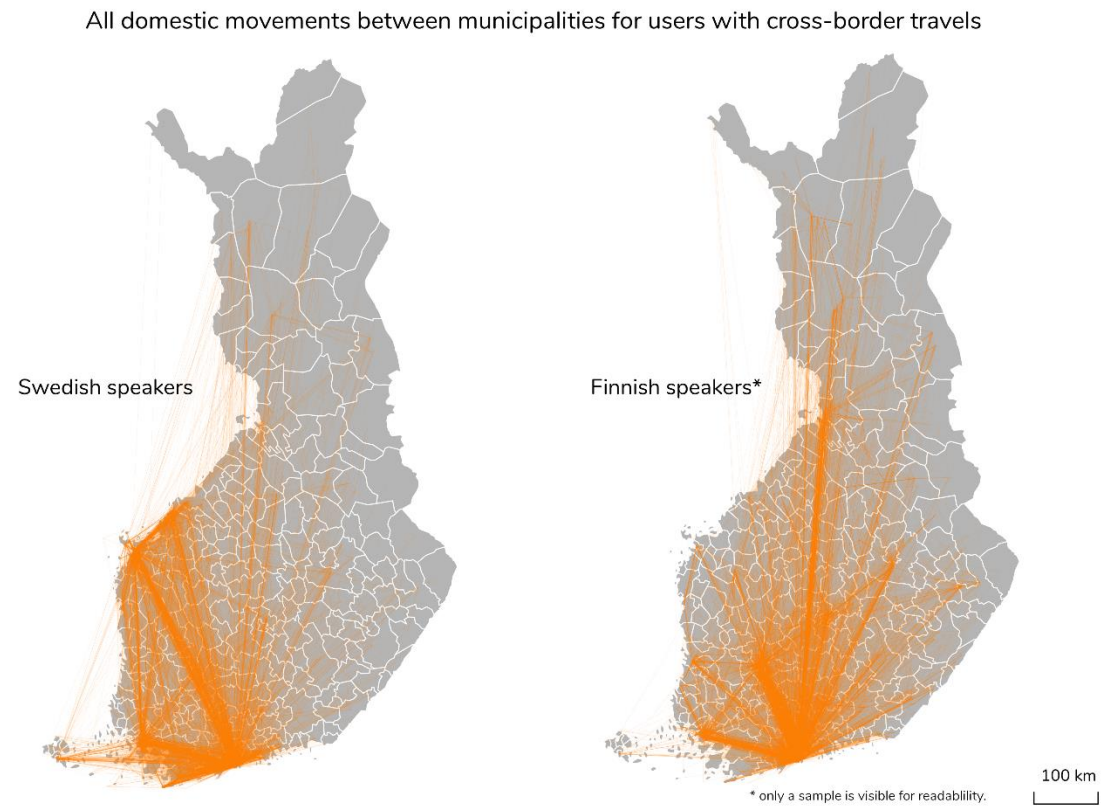

Figure 18 Domestic movements of Swedish and Finnish speakers

5.3 Language Group Comparison

The comparison between groups is made with a one-way ANOVA and adequate post-hoc tests. A significant difference is indicated with a value of less than 0.05. In *Table 9* and *Table 10* we see for each variable where the significant differences are between language groups. For all variables the Estonian language groups results are significantly different from each of the other language groups. This could be a result of not having a large enough sample, and so skewing the data, but it might also be, that there really are significant differences in mobility patterns for Estonian speakers compared to other language groups. Looking at the Shannon entropy and Simpson's index we see that there is a significant difference between Russian speakers and Swedish speakers. For Simpson's index the difference is also significant between Russian and Finnish.

An interesting notion is that there is a lack of significant difference between Swedish and Finnish speakers when looking at the diversity of travels and the number of unique

countries. However, the domestic metrics are significant between Swedish and Finnish, arguing that the mobility patterns within Finland for the two national languages are significantly different. Russian speakers' domestic movements on the other hand, are only significantly different from Estonian speakers. The share of travels to a majority language country is tested for the language minorities and all differences are found to be significant.

Table 9 Post-hoc test results of group comparison on cross-border variables

Significant relationships by the Games-Howell test				
* The difference is significant at the 0.05 level				
Unique Countries				
	Estonian	Russian	Swedish	Finnish
Estonian	NA	0.002*	<0.001*	0.001*
Russian	0.002*	NA	-0.976	-0.748
Swedish	<0.001*	-0.976	NA	0,096
Finnish	0.001*	-0.748	0,096	NA
Shannon Entropy				
	Estonian	Russian	Swedish	Finnish
Estonian	NA	0.005*	<0.001*	<0.001*
Russian	0.005*	NA	0.042	0.074
Swedish	<0.001*	0.042	NA	0,491
Finnish	<0.001*	0.074	0,491	NA
Simpson's Index				
	Estonian	Russian	Swedish	Finnish
Estonian	NA	0.011	<0.001*	<0.001*
Russian	0.011	NA	0.002*	<0.001*
Swedish	<0.001*	0.002*	NA	0,686
Finnish	<0.001*	<0.001*	0,686	NA

Table 10 Post-hoc test results of group comparison on domestic variables and cross-border share

Unique Municipalities				
	Estonian	Russian	Swedish	Finnish
Estonian	NA	0,486	<0.001*	<0.001*
Russian	0,486	NA	<0.001*	<0.001*
Swedish	<0.001*	<0.001*	NA	0,085
Finnish	<0.001*	<0.001*	0,085	NA
Domestic Distance Mean				
	Estonian	Russian	Swedish	Finnish
Estonian	NA	<0.001*	<0.001*	<0.001*
Russian	<0.001*	NA	0,569	0.074
Swedish	<0.001*	0,569	NA	0.010
Finnish	<0.001*	0.074	0.010	NA
Share of travels to majority language country				
	Estonian	Russian	Swedish	
Estonian	NA	<0.001*	<0.001*	
Russian	<0.001*	NA	<0.001*	
Swedish	<0.001*	<0.001*	NA	

5.4 Correlations between Variables

Table 11 Pearson's correlation R-coefficient

Pearson correlation R-coefficient				
Majority lang = Share of travels to majority language country				
Unique Municipalities				
	Unique Countries	Shannon entropy	Simpson Index	Majority lang
Estonian	0.411	0.319	0.293	-0.075
Russian	0.372	0.268	0.216	-0.013
Swedish	0.436	0.396	0.362	0.013
Finnish	0.301	0.290	0.276	NA
Domestic Distance Mean				
	Unique Countries	Shannon entropy	Simpson Index	Majority lang
Estonian	0.162	0.139	0.136	-0.108
Russian	-0.096	-0.088	-0.077	0.074
Swedish	0.007	0.004	0.004	0.010
Finnish	0.005	0.007	0.008	NA

In *Table 11* we can see a summary of Pearson correlations R-coefficient between different variables. The table is divided into two, based on the two different domestic metrics: number of unique municipalities visited, and the domestic mean distance travelled per person. The highest overall correlation is for the Swedish speakers between the number of unique municipalities and the number of unique countries. The correlation is not massive but there seems to be some link between how many unique countries a user visited abroad with the number of unique municipalities visited. In other words, this tells us that a high cross-border mobility often also includes a high domestic mobility. For all language groups the correlation between number of unique countries and number of unique municipalities is the highest correlation out of the used variables. The Shannon entropy correlates less with the unique number of municipalities than compared to the previous variable. Simpson's index even less than Shannon's entropy. This suggests that there is less correlation between how many unique municipalities a user has visited and their travel diversity metrics. Perhaps some other diversity index could give a stronger correlation.

The Share of travels to majority language countries does not have a strong correlation with the number of unique municipalities.

The correlation in general is lower between the domestic distance mean and the selected variables. In fact, it seems like there is very little correlation between all the different variables and the domestic distance mean. This suggests that the distance travelled within Finland is not dependent on the cross-border mobility factors.

6 Discussion

This chapter is divided into three parts, each one discussing one of the research questions. First, I will discuss how well the methods worked. For the most part the analyses seemed to be working for this particular case. There is, however, caveats to be considered. The data is massive, and there are possibilities that unreliable data has been captured as well. The small sample size of Estonian speakers might also endanger how well it works as a proxy for real people. This could be improved by using tweets from a longer time span, but this might in turn pose new issues, like an overgeneralization of the movements. Another data issue would be the number of tweets per user that was used. A very low number of tweets can skew the data towards lower values, but at the same time excluding those might mean, that only very active twitter users are included, and they might not be a good representation of the population.

The way of measuring travel diversity could be further research, as this was just a scratch on the surface. There might be other indices or ways that would explain the diversity of mobility better and this needs to be explored.

Overall, the methods seemed to be functioning for this purpose, but for a more complex and nuanced picture of the cross-border mobility of language minorities we also need more complex tools. I believe the scripts I've created can be of help when developing this further.

6.1 Spatial Connectedness to Neighbouring Country

The first research question posed was "*How are people, belonging to a language minority in Finland spatially connected with neighbouring countries?"*. To answer this, I use geotagged Twitter data, identify the language groups, and then count the number of times each user travelled to the neighbouring country, with a corresponding majority language.

It seems like Estonian Speakers residing in Finland are visiting Estonia more often than the other language groups visit their corresponding neighbouring country. Even though Estonian speakers in Finland tend to be well integrated into Finnish society, depending on their integration dimensions (like work or relationships) they might still have a likely

chance of moving back to Estonia (Anniste & Tammaru, 2014). Most Estonians that move from Estonia to Finland are doing it for work-related reasons and the short distances between the countries have kept ties strong (Kährik & Tammaru, 2019). As the results in this thesis show, the spatial mobility for Estonian speakers is active towards Estonia.

Out of all the cross-border movements made by Estonian, a very high proportion was between Finland and Estonia. A high concentration of movements to and from Estonia might be sign of transnationals that live a very mobile life between these two countries (Ahas et al., 2017; Silm et al., 2020). A study on cross-border movements between Estonia and Finland found that 5 % of all the cross-border travels were made by transnationals which is a significant amount (Silm et al., 2020). In light of this information, it might be worth considering that a part of the Estonian speakers in this study also might be transnationals, considering how concentrated their travels are to and from Estonia.

For Russian speakers the share of travels going to and from Russia was also common on average, but on individual level the differences are substantial. Some Russian speakers have perhaps been fleeing Russia, for example Ingrians, and are therefore not so keen on visiting (Daher et al., 2016). As Russian speakers in Finland are coming from different backgrounds with different reasons for moving, the variation for the relationship with Russia is large (Daher et al., 2016; Mannila & Reuter, 2009). There is of course the proximity of Russia to take into account. It might be possible that some of the movement lines in the east of Finland are the result of cross-border shopping which is popular amongst the Russian middle-class especially (Gurova, 2015; Taipale, 2011).

Swedish speakers travel to Sweden is not as common as for the previous language minorities and their respective major language countries. This could be because Swedish speakers in Finland are an integrated part of Finnish society (Daher et al., 2016). In other words, there are no sense of nationality towards Sweden and perhaps trips are more often conducted for leisure activities. In Finland the movement lines start and end in the traditionally Swedish-speaking regions; the coastal areas of Ostrobothnia as well as the southern coast stretching from Turku to eastern Uusimaa. Young Swedish speakers in Finland have a tendency of moving to Sweden for studies and broader work opportunities, and the trend has been strong in recent years, especially in Ostrobothnia (Kepsu &

Henriksson, 2019). This could explain partially why Stockholm, with important schools and job opportunities is so strongly connected with Swedish speakers. If we look closely, we can also see that the university city Lund, north of Stockholm seems to be a frequently visited place both from southern Finland and Ostrobothnia.

6.2 Spatial Mobility in Finland

For answering the second research question: *"How are people, belonging to a language minority in Finland spatially connected within Finland?"* I use the number of times a user crosses a municipal border and the mean distance for each user. I also look at the number of unique municipalities visited per person and visually inspect the movement lines within Finland.

Estonian speakers seem to have the most sedentary lifestyles when residing in Finland, in relation to the other language groups. Not many unique municipalities are visited per person. This could indicate that many Estonian speakers prefer to spend their weekends and leisure time in Estonia rather than Finland (Silm et al., 2020).

Keeping in mind that the Swedish speaking population is highly concentrated in certain areas it might help explain why the number of unique municipalities per person is less than for Finnish speakers. However, the difference between the number of domestic cross-border movements is very small. This could be an indication that Swedish speakers tend to move about the same amount as Finnish speakers, but that Swedish speakers are usually moving in the Swedish speaking regions, which of course are more spatially limited. A visual inspection of Swedish speakers' movements lines suggests similar patterns.

If we look at the relationship between the cross-border variables and the domestic variables an interesting picture arises. There seems to be some positive correlation between the unique number of countries visited and the unique number of municipalities visited. In other words, a person that has travelled to many unique countries is somewhat likely to also have travelled to many different municipalities. The best correlation was found for Swedish speakers, but all language groups followed a similar pattern.

There seem to be no relevant correlation between the domestic distance mean per user and any of the cross-border variables. This suggests that the domestic distance mean is not

dependent on how many unique countries you visit or the diversity metrics. According to the correlations it might be worth considering, if the domestic distance mean is a variable to be used for these kinds of calculations.

6.3 Differences between Language Groups

When visually inspecting the results within Finland we can see that the spaces where the language groups move differ. This relates to the activity spaces discussed in *2.2*. The spaces differ due to the places where people live. When we look at the language minorities the places people live in becomes visible. The fact that most of the language minorities tend to be more concentrated in certain areas is also reflected by the lower domestic distance mean and perhaps also by the unique municipalities visited.

With the results in chapter *5.4*, we see that there are significant differences between language groups based on multiple variables. An interesting result was the lack of difference between the Swedish and Finnish language groups on the number of unique countries, Shannon entropy and Simpson's index. It means that the two language groups diversity of mobility patterns follow a similar pattern. Perhaps this lack of difference is a result of history and how Swedish speakers are generally perceived as a natural part of Finnish society (Daher et al., 2016).

When discussing travel diversity, it is worth considering the constraints people have for moving around. If look at the coupling constraints presented by Hägerstrand (1970), we can suggest that the Estonian speakers in Finland are constrained by their work being in Finland. Swedish speakers are general quite similar in regards of spatial mobility as Finnish speakers, but we see a slightly smaller domestic distance for Swedish speakers, perhaps due to language constraints, or preference, of being in area, where you can speak your own language. The spatial mobility of Russian speakers seems to be affected by the diversity of the language group.

7 Conclusion

The aim of this thesis was to shed some light on the cross-border movements of different language groups in Finland, based on geotagged Twitter data. The language groups used in this thesis was Estonian, Russian, Swedish, and Finnish. The first three languages are majority in languages in countries bordering Finland. Since we are living in an increasingly mobile world (Sheller & Urry, 2006), we should not neglect the spatial mobility patterns of smaller parts of the population. Some studies show that there are links between ethnic segregation and the size of activity spaces (Järv et al., 2015).

While the number of people belonging to a language minority increases in Finland (Official Statistics of Finland (OSF), 2021) , it is important to also understand their spatial mobility, to help identify differences that might in turn be related to integration into Finnish society.

By using Twitter data, I located the different language minorities and their cross-border mobility patterns as well as their domestic mobility patterns. These helps explain the mobility patterns of the language minority and allow for comparison. To explain the diversity of movements the number of unique countries were calculated, as well as a Shannon entropy and Simpson's Reciprocal index based on that. These measures were compared, as well as the domestic distance travelled, and the number of unique municipalities visited.

The results show that spatial mobility patterns differ between the language groups. Estonian speakers uphold a strong relation to Estonia and visit the country frequently, this might relate to the large number of workers that move across the Finnish bay (Kährik & Tammaru, 2019). Russian speakers' spatial mobility seems to vary more than the other language minorities, which might be due to the diversity within the language group (Shensin, 2008). Swedish speakers are in many regards very similar to Finnish speakers, while they still mostly move from and within the Swedish speaking regions of Finland. In this thesis I have identified differences in spatial mobility between language groups with the help of geo-located big data.

8 Acknowledgements

This thesis has been done in cooperation with the Digital Geography Lab at the university of Helsinki and been linked to the project called BORDERSPACE – Tracing Interactions and Mobilities Beyond State Borders: Towards New Transnational Spaces.

I would like to thank everyone who has helped me sail this ship into harbour. I want to thank my supervisors as well as the kind member of the Digital Geography Lab that has shared their expertise with mer. A special thanks to Håvard Aagesen, without your help and scripts in this thesis would be a lot worse.

Thanks to all wonderful friends who've survived me during this process and provided tips and concrete help with this thesis. If you think "Maybe he is talking about me?", then that means that I mean you. I would like to thank my family, especially my parents and my grandmother. You've all taught me to be curios and the importance of education. I would also like to thank my cat, who does not care about this thesis, but has been happy about me staying at home all the time. Lastly, I want to thank my better half Sini for putting up with me being married to this project. Thank you all for believing in me when I doubted.

9 References

Aagesen, H. W. (2021). *Understanding Functional Cross-border Regions from Twitter Data in the Nordics*. https://helda.helsinki.fi/handle/10138/331462

Ahas, R., Silm, S., & Tiru, M. (2017). Tracking Trans-nationalism with Mobile Telephone Data. In T. Tammaru, R. Eamus, & K. Kallas (Eds.), *Estonian Human Development Report 2016/2017. Estonian at the Age of Migration*. Foundation Estonian Cooperation Assembly.

Anniste, K., & Tammaru, T. (2014). Ethnic differences in integration levels and return migration intentions: A study of Estonian migrants in Finland. *DEMOGRAPHIC RESEARCH*, *30*. https://doi.org/10.4054/DemRes.2013.30.13

Barchiesi, D., Preis, T., Bishop, S., & Moat, H. S. (2015). Modelling human mobility patterns using photographic data shared online. *Royal Society Open Science*, *2*(8). https://doi.org/10.1098/rsos.150046

Blanford, J. I., Huang, Z., Savelyev, A., & MacEachren, A. M. (2015). Geo-located tweets. Enhancing mobility maps and capturing cross-border movement. *PLoS ONE*, *10*(6). https://doi.org/10.1371/journal.pone.0129202

Bourne, R. S. (1916). Trans-National America. *The Atlantic Monthly (1857)*, *118*(1), 86.

Carpentier, S. (2012). Cross-Border Local Mobility between Luxembourg and the Walloon Region: an Overview. *European Journal of Transport and Infrastructure Research*, *12*(2), 198–210.

Crang, P., Dwyer, C., & Jackson, P. (2004). *Transnational Spaces*. Routledge.

Daher, O., Hannikainen, L., & Heikinheimo-Pérez, K. (2016). National minorities in Finland: richness of cultures and languages. In *National minorities in Finland : richness of cultures and languages*. Minority Rights Group Finland.

de Haas, H., & Fokkema, T. (2011). The effects of integration and transnational ties on international return migration intentions. *Demographic Research*, *25*(24), 755–782. https://doi.org/10.4054/DemRes.2011.25.24

Delhey, J., Deutschmann, E., & Cirlanaru, K. (2015). Between ‘class project’ and individualization: The stratification of Europeans’ transnational activities: *Http://Dx.Doi.Org.Libproxy.Helsinki.Fi/10.1177/0268580915578744*, *30*(3), 269–293. https://doi.org/10.1177/0268580915578744

Digital Geography Lab. (2020). *maphel-finlang*. https://github.com/DigitalGeographyLab/maphel-finlang

Ebrahimpour, Z., Wan, W., Luis Velázquez García, J., Cervantes, O., & Hou, L. (2020). Analyzing Social-Geographic Human Mobility Patterns Using Large-Scale Social Media Data. *International Journal of Geo-Information*, *9*(2), 125. https://doi.org/10.3390/ijgi9020125

Faist, T. (2013). The mobility turn: A new paradigm for the social sciences? *Ethnic and Racial Studies*, *36*(11), 1637–1646. https://doi.org/10.1080/01419870.2013.812229

Farber, S., Páez, A., & Morency, C. (2012). Activity spaces and the measurement of clustering and exposure: A case study of linguistic groups in Montreal. *Environment and Planning A*, *44*(2), 315–332. https://doi.org/10.1068/a44203

Feng, K., & Page, S. J. (2010). An Exploratory Study of the Tourism, Migration-Immigration Nexus: Travel Experiences of Chinese Residents in New Zealand. *Current Issues in Tourism*, *3*(3), 246–281. https://doi.org/10.1080/13683500008667875

Finlands grundlag 731/1999. (1999, June 11). https://www.finlex.fi/sv/laki/ajantasa/1999/19990731#a731-1999

Glick Schiller, N., & Salazar, N. B. (2013). Regimes of Mobility Across the Globe. *Journal of Ethnic and Migration Studies*, *39*(2), 183–200. https://doi.org/10.1080/1369183X.2013.723253

Gober, P. (2006). Population, Geography of. In B. Warf (Ed.), *Encyclopedia of Human Geography* (pp. 368–370). Sage publications.

Golledge, R. G., & Stimson, R. J. (1997). *Spatial behavior : a geographic perspective*. Guilford Press.

Gordon, M. M. (1964). *Assimilation in American life : the role of race, religion, and national origins.* Oxford University Press.

Guo, R. (2017). *Cross-Border Resource Management* (3rd ed.). Elsevier. http://ebookcentral.proquest.com/lib/helsinki-ebooks/detail.action?docID=5106120

Gurova, O. (2015). From Russia to Finland. In *Fashion and the Consumer Revolution in Contemporary Russia* (1st ed., pp. 116–132). Routledge. http://www.russiatourism.ru/data/File/Statistika/2011/viezd_2010_k_2009.xls

Haddawy, P., Lawpoolsri, S., Sa-ngamuang, C., Su Yin, M., Barkowsky, T., Wiratsudakul, A., Kaewkungwal, J., Khamsiriwatchara, A., Sa-angchai, P., Sattabongkot, J., & Cui, L. (2021). Effects of COVID-19 government travel restrictions on mobility in a rural border area of Northern Thailand: A mobile phone tracking study. *PLOS ONE*, *16*(2), e0245842. https://doi.org/10.1371/journal.pone.0245842

Hägerstrand, T. (1970). What about people in Regional Science? *Papers of the Regional Science Association*, *24*, 6–21. https://doi.org/https://doi.org/10.1007/BF01936872

Hasanzadeh, K., Czepkiewicz, M., Heinonen, J., Kyttä, M., Ala-Mantila, S., & Ottelin, J. (2019). Beyond geometries of activity spaces: A holistic study of daily travel patterns, individual characteristics, and perceived wellbeing in helsinki metropolitan area. *Journal of Transport and Land Use*, *12*(1), 149–177. https://doi.org/10.5198/jtlu.2019.1148

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., & Ratti, C. (2014). Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, *41*(3), 260–271. https://doi.org/10.1080/15230406.2014.890072

Heikinheimo, V., Tenkanen, H., Hiippala, T., & Toivonen, T. (2018). Digital Imaginations of National Parks in Different Social Media: A Data Exploration. In R. Westerholt, F.-B. Mocnik, & A. Zipf (Eds.), *First Workshop on Platial Analysis (PLATIAL'18)* (pp. 45–52). https://doi.org/10.5281/zenodo.1472745

Hiippala, T., Väisänen, T., Toivonen, T., & Järv, O. (2020). Mapping the languages of Twitter in Finland: Richness and diversity in space and time. *Neuphilologische Mitteilungen*, *121*(1), 12–44. https://journal.fi/nm/article/view/99996

Horton, F., & Reynolds, D. R. (1970). Action Space Formation: A Behavioral Approach to Predicting Urban Travel Behavior. *Highway Research Record*, *322*, 136–148.

IBM Corp. (2019). *IBM SPSS Statistics for Windows*. https://hadoop.apache.org

International Organization for Migration. (2021). *Key Migration Terms*. https://www.iom.int/key-migration-terms

Ipatti, L. (2019). AT THE ROOTS OF THE "FINLAND BOOM." *Scandinavian Journal of History*, *44*(1), 103–130. https://doi.org/10.1080/03468755.2018.1502680

Jakle, J. A., Brunn, Stanley., & Roseman, C. C. (1976). *Human spatial behavior : a social geography*. Duxbury Pr.

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

Järv, O., Müürisepp, K., Ahas, R., Derudder, B., & Witlox, F. (2015). Ethnic differences in activity spaces as a characteristic of segregation: A study based on mobile phone usage in Tallinn, Estonia. *Urban Studies*, *52*(14), 2680–2698. https://doi.org/10.1177/0042098014550459

Järv, O., Tominga, A., Müürisepp, K., & Silm, S. (2021). The impact of COVID-19 on daily lives of transnational people based on smartphone data: Estonians in Finland. *Journal of Location Based Services*, *15*(3), 169–197. https://doi.org/10.1080/17489725.2021.1887526

Kährik, A., & Tammaru, T. (2019). *Virolaiset Suomessa*. Elo. https://www.tuglas.fi/index.php?id=237

Kaufmann, V. (2003). *Re-Thinking Mobility: Contemporary Sociology*. Routledge. https://doi.org/10.4324/9781315244303

Kaufmann, V., Bergman, M. M., & Joye, D. (2004). Motility: Mobility as Capital. *International Journal of Urban and Regional Research*, *28*(4), 745–756. https://doi.org/10.1111/j.0309-1317.2004.00549.x

Kaufmann, V., Dubois, Y., & Ravalet, E. (2018). Measuring and typifying mobility using motility. *Applied Mobilities*, *3*(2), 198–213. https://doi.org/10.1080/23800127.2017.1364540

Kellerman, A. (2012). *Daily Spatial Mobilities: Physical and Virtual* (1st ed.). Milton: Taylor & Francis. https://doi.org/10.4324/9781315575780

Kepsu, K., & Henriksson, B. (2019). *Hjärnflykt eller inte? Del II. Svenskspråkig ungdomsflyttning till Sverige: trender och drivkrafter*. Tankesmedjan Magma. www.magma.fi

Kitchin, R. (2014). *The data revolution : big data, open data, data infrastructures and their consequences*. SAGE. https://doi.org/10.4135/9781473909472

Kotimaisten kielten keskus. (2021). *Kielet*. https://www.kotus.fi/kielitieto/kielet

Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety*.

Lemoine, N. (2015). *Species Diversity*. https://ecopy.readthedocs.io/en/latest/diversity.html

Levitt, P., & Jaworsky, B. N. (2007). Transnational Migration Studies: Past Developments and Future Trends. *Annual Review of Sociology*, *33*(1), 129–156. https://doi.org/10.1146/annurev.soc.33.040406.131816

Levitt, P., & Schiller, N. G. (2004). Conceptualizing Simultaneity: A Transnational Social Field Perspective on Society. *The International Migration Review*, *38*(3), 1002–1039.

Liu, Q., Wang, Z., & Ye, | Xinyue. (2018). Comparing mobility patterns between residents and visitors using geo-tagged social media data. *Transactions in GIS*, *22*, 1372–1389. https://doi.org/10.1111/tgis.12478

Lynch, K. (1960). The image of the city. In *The image of the city*. Technology Press.

Ma, Z. (2011). Hägerstrand time-geography revisited from current perspective. *2011 International Conference on Multimedia Technology, ICMT 2011*, 5407–5410. https://doi.org/10.1109/ICMT.2011.6002703

Mannila, S., & Reuter, A. (2009). Social Exclusion Risks and their Accumulation among Russian-Speaking, Ethnically Finnish and Estonian Immigrants to Finland. *Journal of Ethnic and Migration Studies*, *35*(6), 939–956. https://doi.org/10.1080/13691830902957718

Massey, D. S., & Denton, N. A. (1988). The Dimensions of Residential Segregation. *Social Forces*, *67*(2), 281. https://doi.org/10.2307/2579183

Massinen, S. (2019). *Modeling Cross-Border Mobility Using Geotagged Twitter in the Greater Region of Luxembourg*. . https: //helda.helsinki.fi/handle/10138/306530

Mazzoni, D., Albanesi, C., Ferreira, P. D., Opermann, S., Pavlopoulos, V., & Cicognani, E. (2017). Cross-border mobility, European identity and participation among European adolescents and young adults. *European Journal of Development Psychology*, *15*(3), 324–339. https://doi.org/10.1080/17405629.2017.1378089

Merriam-Webster Inc. (n.d.-a). *mobility*. Merriam-Webster.Com. Retrieved March 12, 2021, from https://www.merriam-webster.com/dictionary/mobility

Merriam-Webster Inc. (n.d.-b). *transnational*. Merriam-Webster.Com. Retrieved February 1, 2021, from https://www.merriam-webster.com/dictionary/transnational

Miller, H. J. (2005). A Measurement Theory for Time Geography. *Geographical Analysis*, *37*, 17–45. https://doi.org/10.1111/j.1538-4632.2005.00575.x

Minghi, J. v. (1963). Boundary Studies in Political Geography. *Annals of the Association of American Geographers*, *53*(3), 407–428.

Mooses, V., Silm, S., Tammaru, T., & Saluveer, E. (2020). An ethno-linguistic dimension in transnational activity space measured with mobile phone data. *Humanities and Social Sciences Communications*, *7*(1). https://doi.org/10.1057/s41599-020-00627-3

Nonnenmacher, L., Baumann, M., le Bihan, E., Askenazy, P., & Chauvel, L. (2021). Cross-border mobility in European countries: associations between cross-border worker status and health outcomes. *BMC Public Health*, *21*(588). https://doi.org/10.1186/s12889-021-10564-8

Official Statistics Finland (OSF). (2020). *Population 31.12. by Area, Sex, Year, Language and Information.*

Official Statistics of Finland (OSF). (2021). *Population structure [e-publication]*. http://www.stat.fi/til/vaerak/index_en.html

Paasi, A., & Prokkola, E.-K. (2008). Territorial Dynamics, Cross-border Work and Everyday Life in the Finnish-Swedish Border Area. *Space and Polity*, *12*(1), 13–29. https://doi.org/10.1080/13562570801969366

Prindiville, N., & Hjelm, T. (2018). The "secularization" and ethnicization of migration discourse: the Ingrian Finnish Right to Return in Finnish politics. *Ethnic and Racial Studies*, *41*(9), 1574–1593. https://doi.org/10.1080/01419870.2017.1312011

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

Shannon, C. E. (1949). The mathematical theory of communication . In *The mathematical theory of communication*. University of Illinois Press.

Shaw, S. L., & Yu, H. (2009). A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical-virtual space. *Journal of Transport Geography*, *17*(2), 141–149. https://doi.org/10.1016/j.jtrangeo.2008.11.012

Sheller, M., & Urry, J. (2006). The new mobilities paradigm. *Environment and Planning A*, *38*(2), 207–226. https://doi.org/10.1068/a37268

Sheller, M., & Urry, J. (2016). Mobilizing the new mobilities paradigm. *Applied Mobilities*, *1*(1), 10–25. https://doi.org/10.1080/23800127.2016.1151216

Shensin, V. (2008). *VENÄLÄISET JA VENÄLÄINEN KULTTUURI SUOMESSA*.

Shirazi, M. R. (2018). *Mapping neighbourhood outdoor activities: space, time, gender and age*. https://doi.org/10.1080/13574809.2018.1458607

Silm, S., & Ahas, R. (2014). The temporal variation of ethnic segregation in a city: Evidence from a mobile phone use dataset. *Social Science Research*, *47*, 30–43. https://doi.org/10.1016/J.SSRESEARCH.2014.03.011

Silm, S., Ahas, R., & Mooses, V. (2017). Are younger age groups less segregated? Measuring ethnic segregation in activity spaces using mobile phone data. *Journal of Ethnic and Migration Studies*, *44*(11), 1797–1817. https://doi.org/10.1080/1369183X.2017.1400425

Silm, S., Jauhiainen, J. S., Raun, J., & Tiru, M. (2020). Temporary population mobilities between Estonia and Finland based on mobile phone data and the emergence of a cross-border region. *European Planning Studies*, 1–21. https://doi.org/10.1080/09654313.2020.1774514

Simpson, E. (1949). Measurement of Diversity. *Nature 1949 163:4148*, *163*(4148), 688–688. https://doi.org/10.1038/163688a0

Sorokin, P. A. (1941). Social and cultural dynamics. 4, Basic problems, principles and methods . In *Social and cultural dynamics. 4, Basic problems, principles and methods*.

Spellerberg, I. F., & Fedor, P. J. (2003). A tribute to Claude Shannon (1916-2001) and a plea for more rigorous use of species richness, species diversity and the “Shannon-Wiener” Index. *Global Ecology & Biogeography*, *12*, 177–179.

Taipale, H. (2011). *TAX FREE-ostosen määrän vaikutus tullin toimintaan*. https://www.theseus.fi/bitstream/handle/10024/32664/Taipale_Heini.pdf?sequence=1

Toomet, O., Silm, S., Saluveer, E., Ahas, R., & Tammaru, T. (2015). *Where Do Ethno-Linguistic Groups Meet? How Copresence during Free-Time Is Related to Copresence at Home and at Work*. https://doi.org/10.1371/journal.pone.0126093

Twitter. (2021). *Twitter API*. https://developer.twitter.com/en/docs/twitter-api

Väisänen, T., Sirkiä, S., Hiippala, T., Järv, O., & Toivonen, T. (2021). *DigitalGeographyLab/tweetsearcher: A Python tool for downloading Tweets using Academic Research credentials*. Zenodo. https://doi.org/10.5281/zenodo.4767170

van der Velde, M., & van Naerssen, T. (2010). People, borders, trajectories: an approach to cross-border mobility and immobility in and to the European Union. *AREA*, *43*(2), 218–224. https://doi.org/10.1111/j.1475-4762.2010.00974.x

van Houtum, H. (2005). The geopolitics of borders and boundaries. In *Geopolitics* (Vol. 10, Issue 4, pp. 672–679). https://doi.org/10.1080/14650040500318522

Vanderbeck, R. (2006). Segregation. In B. Warf (Ed.), *Encyclopedia of Human Geography* (pp. 423–424). Sage Publications.

Vertovec, S. (2001). Transnationalism and identity. *Journal of Ethnic and Migration Studies*, *27*(4), 573–582. https://doi.org/10.1080/13691830120090386

Vertovec, S. (2009). *Transnationalism*. Routledge.

Verwiebe, R., Reinprecht, C., Haindorfer, R., & Wiesboeck, L. (2017). *How to Succeed in a Transnational Labor Market: Job Search and Wages among Hungarian, Slovak, and Czech Commuters in Austria*. *51*, 251. https://doi.org/10.1111/imre.12193

Walton-Roberts, M. (2004). Transnational Migration Theory in Population Geography: Gendered Practices in Networks Linking Canada and India. *Population, Space and Place*, *10*(5), 361–373. https://doi.org/10.1002/psp.343

Wong, D. W. S., Shaw, S.-L., Wong, D. W. S., & Shaw, S.-L. (2011). Measuring segregation: an activity space approach. *Journal of Geographical Systems*, *13*, 127–145. https://doi.org/10.1007/s10109-010-0112-x

Zhang, X., Wang, J., Kwan, M.-P., & Chai, Y. (2019). Reside nearby, behave apart? Activity-space-based segregation among residents of various types of housing in Beijing, China. *Cities*, *88*, 166–180. https://doi.org/10.1016/j.cities.2018.10.009

Appendix A

Movement lines on a global scale for Estonian, Swedish and Finnish speakers residing in Finland.

Estonian speakers transnational movements globally

Swedish speakers transnational movements globally

Finnish speakers transnational movements globally

Appendix B

Movement lines on a neighbouring country scale for Russian and Swedish speakers.

Swedish speakers movements to neighbouring countries