Studying Segregation In Estonia Using Call Data Records - Supplementary File

July 12, 2021

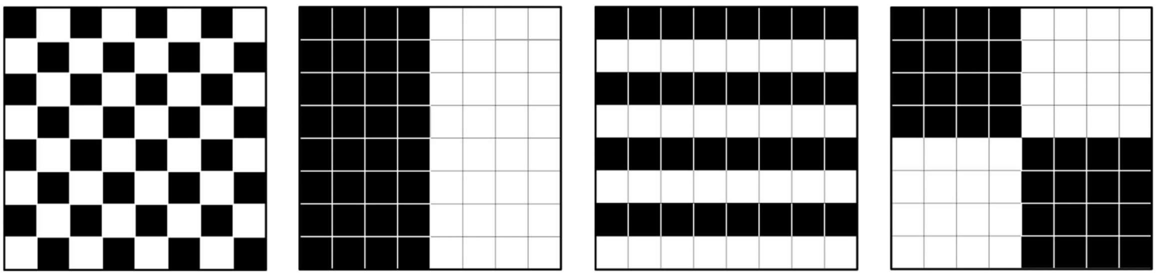
Contents

1 Related Work

The Table 1 shows various non-spatial and spatial indices. The 'Checkerboard problem' is an example that is typically used to investigate the properties of various segregation measures [29, 30, 7, 14] is shown in Figure 1. In most people's perceptions, the four arrangements reflect varying levels of segregation. However, the values of D are the same, i.e., it does not differ between different spatial arrangements. In fact, the value of D will still be 1 as long as one population group is inhabited exclusively in each spatial unit [25, 20, 27].

Non-Spatial Indices			**Spatial Indices**		
S.No.	**Index**	**Citation**	**S.No.**	**Index**	**Citation**
1	Dissimilarity index	[8]	11	Spatial proximity (SP) index	[25]
2	Gini index	[8]	12	Multi-group SP	[13]
3	Centralisation index	[9]	13	Dissimilarity index incorporating spatial adjacency	[20]
4	Coleman's Homophily index	[6]	14	Dssimilarity index incorporating common boundary lengths	[27]
5	Freeman's index	[11]	15	Dssimilarity index incorporating common boundary lengths and perimeter/area ratio	[27, 23]
6	Multi–group dissimilarity index	[19, 24]	16	Spatial version of multigroup dissimilarity index	[28, 15]
7	Exposure index	[17]	17	General index of spatial segregation	[26]
8	Neighbourhood sorting index (NSI)	[16]	18	Generalised spatial dissimilarity (GSD) index	[10]
9	Typology for classifying ethnic residential areas	[22, 21]	19	Spatial dissimilarity index	[23]
10	Location Quotient (LQ)	[4]	20	Spatial information theory index	[23]

Table 1: Non-spatial and spatial segregation indices.



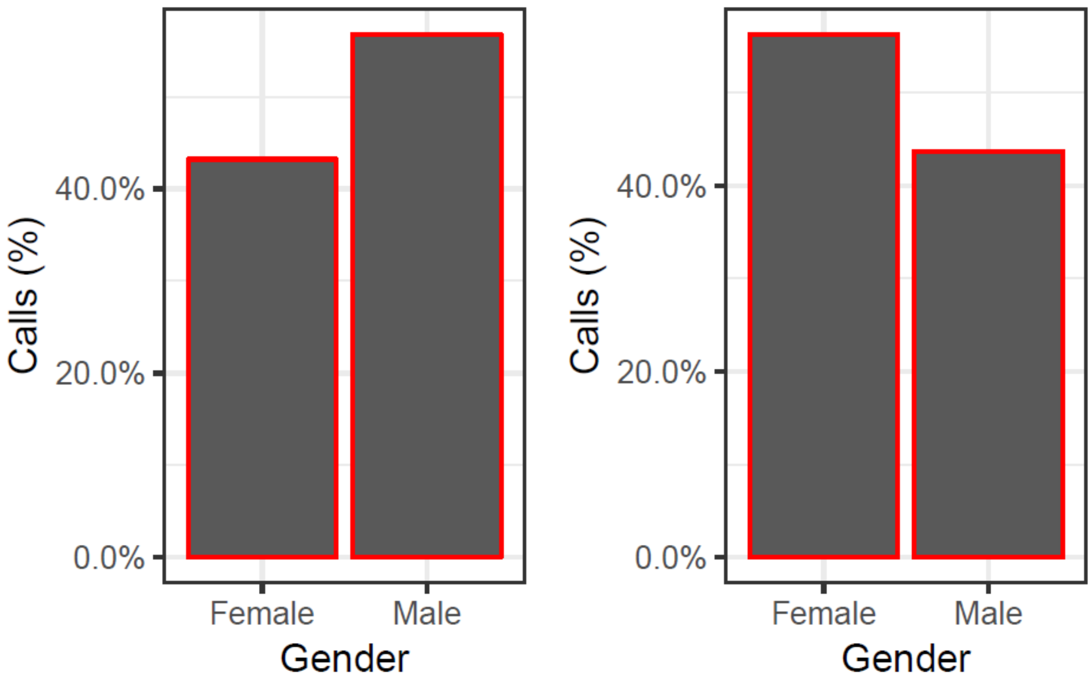
(a) (b) (c) (d)

Figure 1: The Checkerboard Problem.

2 Descriptive Analysis

2.1 Gender Segregation

We begin our analysis by analyzing the percentage of calls that remain within the same gender against the opposite gender (see Figure 2). We observe that 56.77% male calls to male and 56.28% female calls remain within females indicating that the same gender calls happen more. Next, we perform the t-test, and the majority percentage of within-gender calls and the t-test's p-value of 2.2e-16 suggests that the users prefer to call within the same gender.



(a) Calls initiated by *males* (b) Calls initiated by *females*

Figure 2: **Calls initiated by males and females.** blueHere, x-axis represents gender of the called person and y-axis represents the percentage of calls. Figure (a) shows the calls initiated by *males* to *females* and *males*. Similarly, figure (b) shows the calls initiated by *females*.

2.2 Age-Group Segregation

Figure 3 shows the *FSI* values for different age-groups in various Estonian counties. The red solid horizontal line shows the average *FSI* index based on age-group segregation. The *FSI* index for various age-group is also shown using different color bars. We can observe that *FSI* index for all age-groups is less than the average *FSI* index in *Voru*, *Polva*, *Saare* and *Rapla*. We can also observe that in *Jogeva*, *Laane* and *Laane-Viru*, the segregation index for age-group (24-54) is approximately near to the average segregation. The interesting observation to be noted here is that the age-group (24-54) is mostly segregated in county *Parnu*. Another interesting observation can be that in *Hiiu* county (highlighted using a blue rectangle in Figure 3), only elderly (i.e., (64,100)) individuals are more segregated than the mean *FSI* value. Therefore, we further analyzed this age-group for Hiiu county using social network analysis.

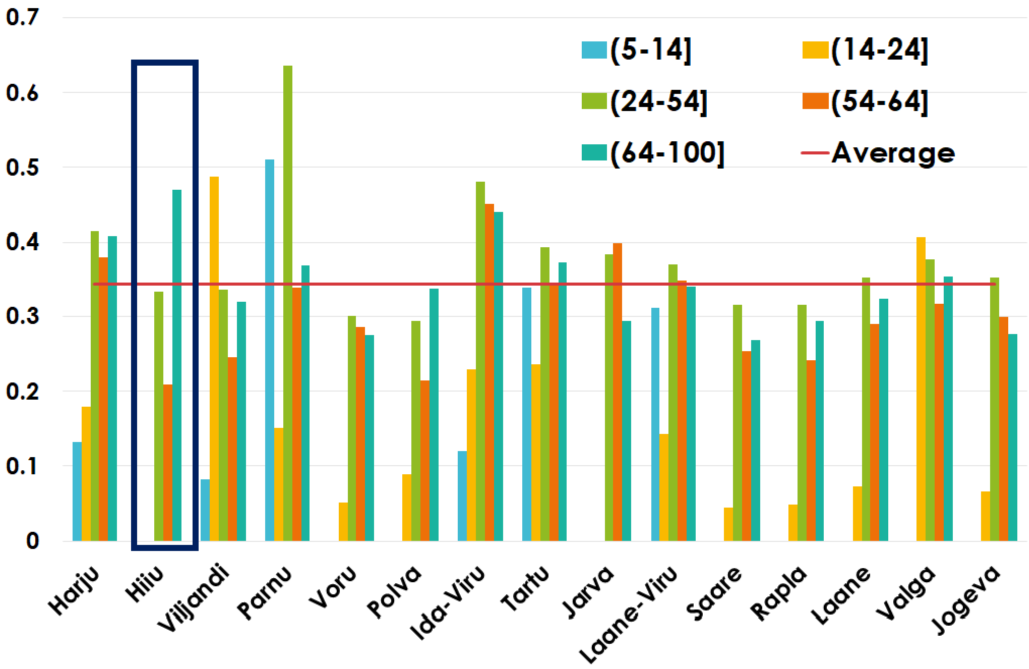
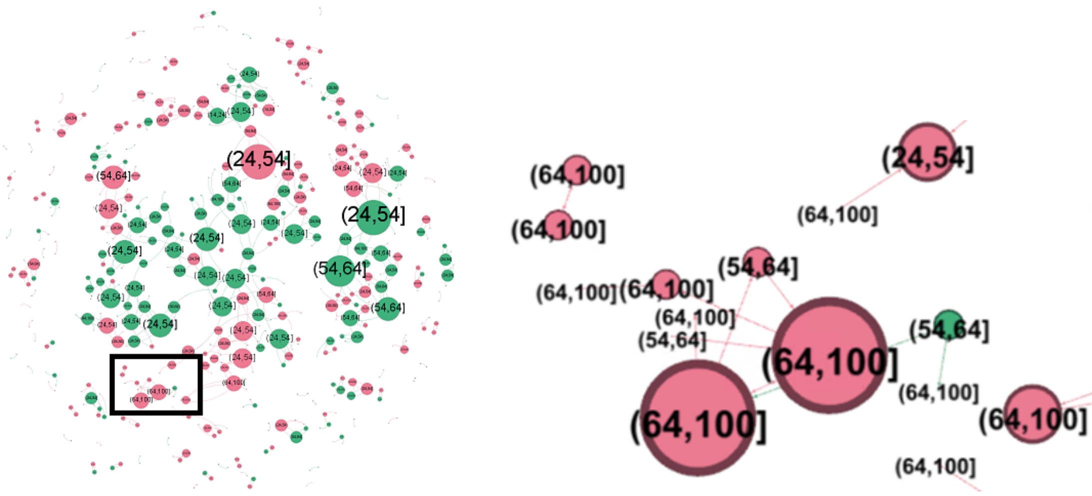


Figure 3: The *FSI* values for age-groups segregation in Estonia counties.

We create a directed network for *Hiiu* county that represents the call connections among users where an edge $(u \longrightarrow v)$ is formed if a user u has called user v. Fig 4a shows the CDR network, where each node is color-coded based on the gender. The red nodes represent male users, and green nodes represent female users. Links between users are also color-coded. Links which originate from males are colored red (i.e., calls from male to male; and male to a female), and similarly links that originate from females are colored green (i.e., calls from female to female; and female to male). The node labels indicate the age group to which the user belongs.

Furthermore, we employ the well-known *degree centrality* algorithm [12] to identify the centrality of nodes in the network. Degree centrality is the most basic metric for determining node influence. A node's impact grows as its connections increase. The normalized degree centrality is used to



(a) Age-group interaction in Hiiu

(b) Network subset

Figure 4: **Users network in Hiiu using CDR data.** The original network of users in Hiiu county of Estonia. In Figure (a), users are color-coded based on *gender*. Color coding is as follows: red node are male users and green nodes are female users. The nodes with higher degree value are shown in relatively bigger size than others. Node labels are their age-group. Figure (b) shows a specific portion of the network (highlighted using rectangle in Figure (a)).

compare the influences of nodes in the network and is defined as follows:

$$DC(i) = \frac{d_i}{n-1} \tag{1}$$

where, $n = |V|$ is the number of nodes in our network and $n-1$ is the largest possible degree. An individual with a higher degree centrality could reflect its greater social influence to propagate a piece of information in the network. In Fig 4a, the size of the node reflects the degree centrality of the node.

In Figure 4b, we shown a subset of the network highlighted using rectangle in Figure 4a. Although highlighting a network subset can be called biased such that we may be hiding the other subsets where similar relationships may be missing, nevertheless it can provide some significant information about segregation of elderly (i.e., (64,100)) individuals in the actual network. We can observe that elderly individuals prefer to connect other elderly individuals. We also observe that there are many elderly pairs which are only connected to each other.

2.3 Segregation in Estonian Counties

2.3.1 Segregation in Estonian counties without Narva city

We measured the *FSI* values for each county after removing the *Narva* city of *Ida-Viru* county. The reason for removing this particular city is the fact that majority of its population has Russian-speaking individuals. In particular, 95.7% of the population of *Narva* city are native Russian-speakers, and 87.7% are ethnic Russians [2]. We calculated *FSI* index after removing *Narva* city and find that the segregation of *Ida-Viru* county decreased from 0.856 to 0.843 as shown in Figure 5 using green color rectangle. Hence, we can conclude that Russian-speaking population in *Narva* city of *Ida-Viru* county is more segregated than Estonian-speaking population. We can further infer that location segregation and language segregation are correlated with each other.

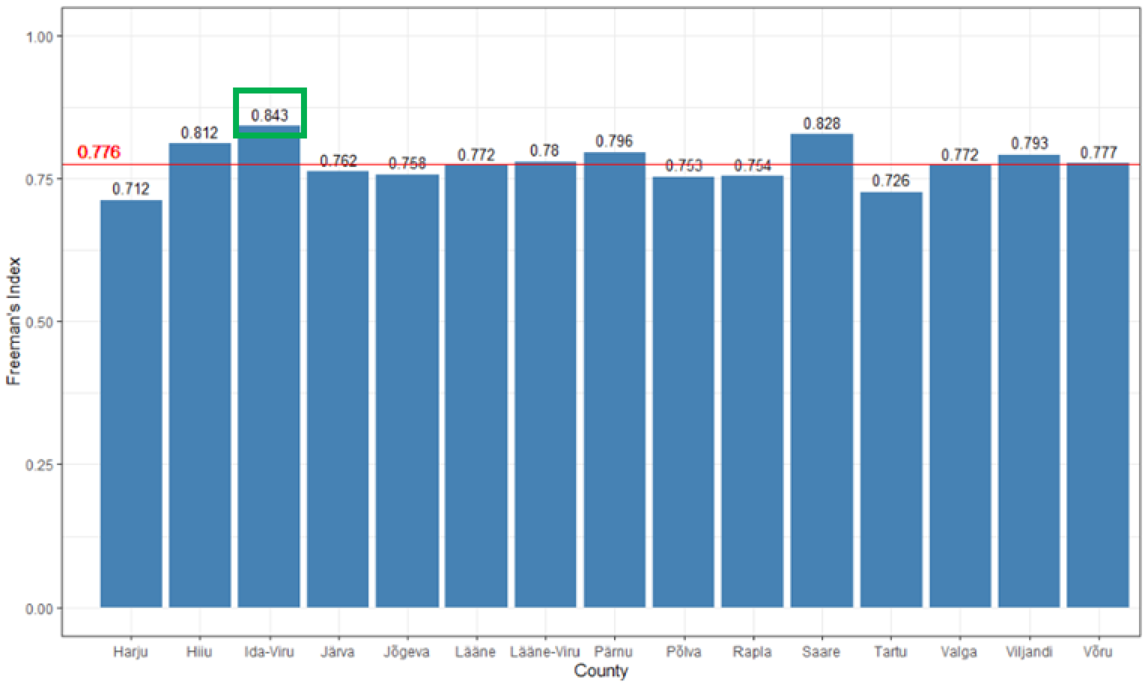


Figure 5: The *FSI* values for location segregation in Estonia after removing *Narva* city of *Ida-Viru* county.

2.3.2 Segregation in Estonian counties without Pärnu city

In this section, we measured the *FSI* values for each county after removing the *Pärnu* city of *Pärnu* county. The reason for removing this particular city is the fact that majority of its population has Estonian-speaking individuals (83%) [2]. On the other hand, the Russian-speaking, Ukrainian-speaking and other language speaking individuals are 12.8%, 1.7% and 2.5% respectively [2]. We calculated *FSI* index after removing *Pärnu* city and find that the segregation of *Pärnu* and *Haju* county increased from 0.712 and 0.796 to 0.718 and 0.805 respectively as shown in Figure 6 using red color rectangles. Hence, we can conclude that Estonian-speaking population in *Pärnu* city of *Pärnu* county is less segregated than other language speaking population. In other words, we see the similar behavior as earlier that Russian-speaking population in *Pärnu* county is more segregated. Therefore, we can infer with reasonable confidence that location segregation and language segregation are correlated with each other.

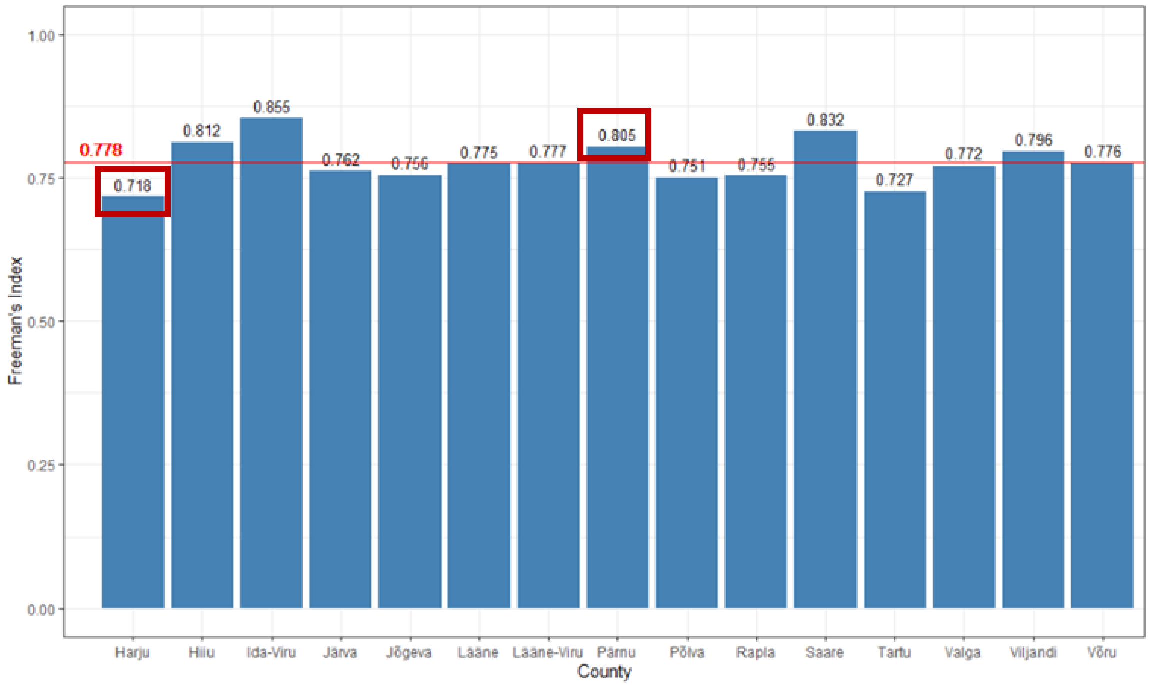


Figure 6: The *FSI* values for location segregation in Estonia after removing *Pärnu* city of *Pärnu* county.

2.3.3 Segregation in Estonian counties without Tartu city

Here, we measured the *FSI* values for each county after removing the *Tartu* city of *Tartu* county. The reason for removing this particular city is the fact that it is the second largest city of Estonia, after Estonia's political and financial capital Tallinn. *Tartu* is often considered the intellectual centre of the country [1, 3, 5], especially since it is home to the nation's oldest and most renowned university, the University of Tartu. *Tartu* city itself is also the oldest city of Estonia [18].

The majority population of *Tartu* city speaks Estonian (80%) [2]. On the other hand, the Russian-speaking and other language speaking individuals are 14% and 6% respectively [2]. We calculated *FSI* index after removing *Tartu* city and find that it affects many counties. The *FSI* value for six counties (*Harju*, *Jõgeva*, *Põlva*, *Tartu*, *Valga* and *Võru*) have increased as shown in Figure 7 using red color rectangles. On the other hand, the *FSI* value for four counties (*Hiiu*, *Jõgeva*, *Rapla* and *Saare*) have decreased as shown in Figure 7 using green color rectangles. Hence, we can conclude that individuals in *Tartu* city of *Tartu* county have greater impact on other counties either by increasing or decreasing their *FSI* values. Interestingly, we also observe that removing *Tartu* city has no impact on *Ida-Viru* and *Pärnu* counties which shows that their segregation is independent of *Tartu* city. This is also true for other counties which are not affected by removal of *Tartu* city. Therefore, we can conclude that *Tartu* city is well connected with other counties of Estonia.

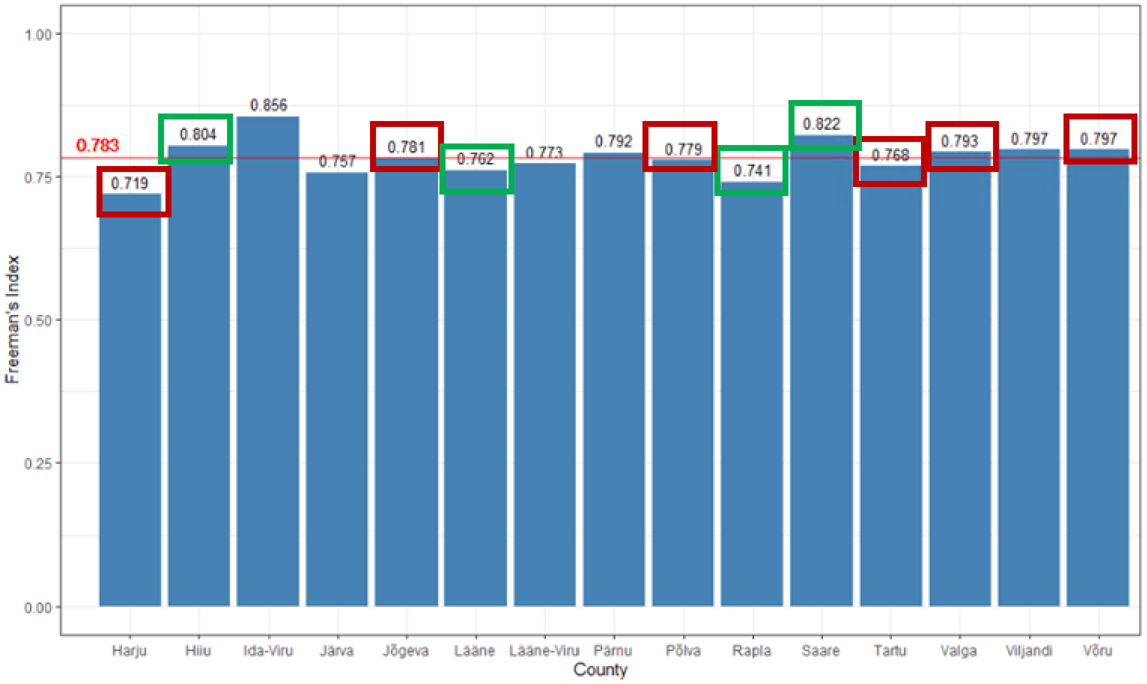


Figure 7: The *FSI* values for location segregation in Estonia after removing *Tartu* city of *Tartu* county.

2.3.4 Segregation in Estonian counties without Tallinn city

Next, we measure the *FSI* values for each county after removing the *Tallinn* city of *Harju* county. The reason for removing this particular city is that it is the capital and the most populous city of Estonia. *Tallinn* is also the main financial, industrial and cultural centre of Estonia [1]. *Tallinn* city also have approximately equal number of Estonian-speaking (50.1%) and Russian- speaking (46.7%) individuals [2].

We calculated *FSI* index after removing *Tallinn* city and find that it affects all counties. The *FSI* value for eight counties (*Harju*, *Hiiu*, *Ida-Viru*, *Lääne*, *Lääne-Viru*, *Pärnu*, *Rapla* and *Saare*) have increased as shown in Figure 8 using red color rectangles. Hence, we can conclude that these eight counties are very well-connected to the *Tallinn* city individuals. On the other hand, the *FSI* value for seven counties (*Järva*, *Jõgeva*, *Põlva*, *Tartu*, *Valga*, *Viljandi* and *Võru*) have decreased as shown in Figure 8 using green color rectangles. This indicate that these counties are well-connected with other cities and counties other than *Tallinn* city.

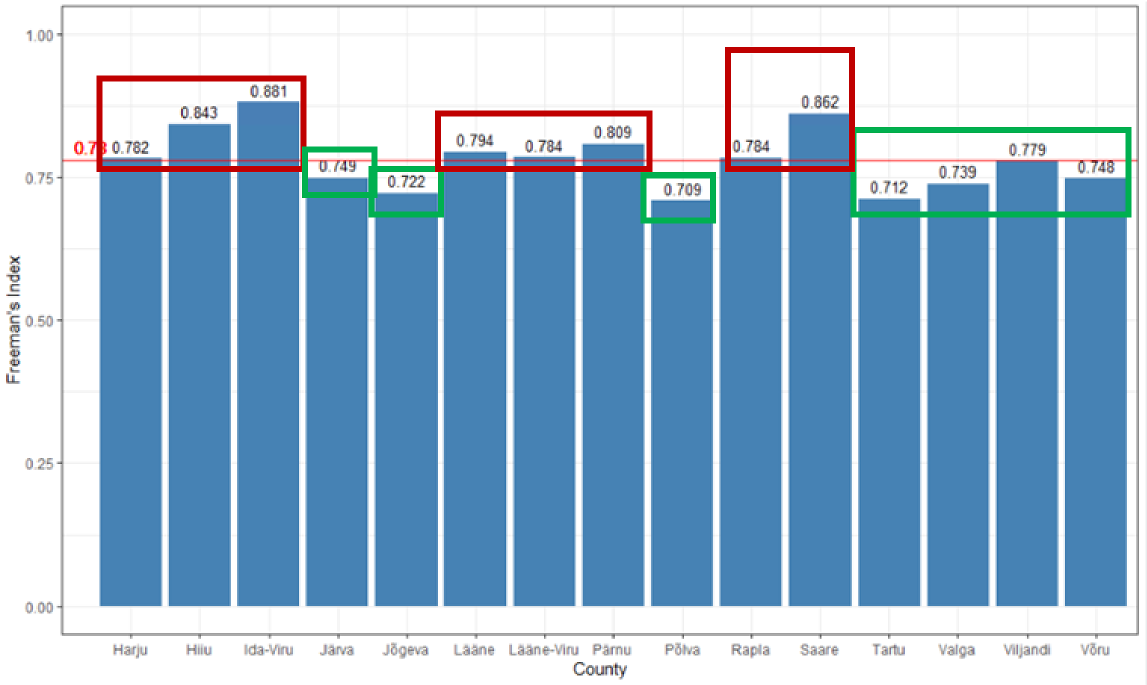


Figure 8: The *FSI* values for location segregation in Estonia after removing *Tallinn* city of *Harju* county.

References

[1] Nordisch.info. https://www.nordisch.info/estland/. Accessed: 2021-03-20.

[2] Statistics estonia. https://andmed.stat.ee/et/stat. Accessed: 2021-03-20.

[3] Jonathan Bousfield. *The Rough Guide to Estonia, Latvia & Lithuania*. Dorling Kindersley Ltd, 2011.

[4] Lawrence A Brown and Su-Yeul Chung. Spatial segregation, segregation indices and the geographical perspective. *Population, space and place*, 12(2), 2006.

[5] Sergey Chernov. Tartu: Estonia's intellectual and theater capital. https://www.themoscowtimes.com/2012/12/23/tartu-estonias-intellectual-and-theater-capital-a20396. Published: Dec. 24 2012.

[6] James Coleman. Relational analysis: The study of social organizations with survey methods. *Human organization*, 17(4):28–36, 1958.

[7] Casey J Dawkins. Measuring the spatial pattern of residential segregation. *Urban Studies*, 41(4):833–851, 2004.

[8] Otis Dudley Duncan and Beverly Duncan. A methodological analysis of segregation indexes. *American sociological review*, 20(2):210–217, 1955.

[9] Otis Dudley Duncan and Beverly Duncan. Residential distribution and occupational stratification. *American journal of sociology*, 60(5):493–503, 1955.

[10] Flávia F Feitosa, Gilberto Camara, Antônio Miguel Vieira Monteiro, Thomas Koschitzki, and Marcelino PS Silva. Global and local spatial indices of urban segregation. *International Journal of Geographical Information Science*, 21(3):299–323, 2007.

[11] Linton C Freeman. Segregation in social networks. *Sociological Methods & Research*, 6(4):411–429, 1978.

[12] Rahul Goel, Anurag Singh, and Fakhteh Ghanbarnejad. Modeling competitive marketing strategies in social networks. *Physica A: Statistical Mechanics and its Applications*, 518:50–70, 2019.

[13] Rick Grannis. Discussion: Segregation indices and their functional inputs. *Sociological methodology*, 32(1):69–84, 2002.

[14] Richard Harris. Measuring segregation as a spatial optimisation problem, revisited: a case study of london, 1991–2011. *International Journal of Geographical Information Science*, 30(3):474–493, 2016.

[15] A Horn. Measuring multi-ethnic spatial segregation in south african cities. *South African Geographical Journal*, 87(1):58–72, 2005.

[16] Paul A Jargowsky. Take the money and run: Economic segregation in us metropolitan areas. *American sociological review*, pages 984–998, 1996.

[17] Stanley Lieberson. An asymmetrical approach to segregation. *Ethnic segregation in cities.*, pages 61–82, 1981.

[18] M Mets and Renita Raudsepp. *Baltic Piling.* CRC Press, 2013.

[19] Barrie S Morgan. The segregation of socio-economic groups in urban areas: a comparative analysis. *Urban Studies*, 12(1):47–60, 1975.

[20] Richard Morrill. On the measure of geographic segregation. In *Geography research forum*, volume 11, pages 25–36, 1991.

[21] Michael Poulsen, Ron Johnson, and James Forrest. Plural cities and ethnic enclaves: introducing a measurement procedure for comparative study. *International Journal of Urban and Regional Research*, 26(2):229–243, 2002.

[22] Michael Poulsen, Ron Johnston, and James Forrest. Intraurban ethnic enclaves: introducing a knowledge-based classification method. *Environment and planning A*, 33(11):2071–2082, 2001.

[23] Sean F Reardon and David O'Sullivan. Measures of spatial segregation. *Sociological methodology*, 34(1):121–162, 2004.

[24] James M Sakoda. A generalized index of dissimilarity. *Demography*, 18(2):245–250, 1981.

[25] Michael J White. The measurement of spatial segregation. *American journal of sociology*, 88(5):1008–1018, 1983.

[26] David W Wong. Formulating a general spatial segregation measure. *The Professional Geographer*, 57(2):285–294, 2005.

[27] David WS Wong. Spatial indices of segregation. *Urban studies*, 30(3), 1993.

[28] David WS Wong. Measuring multiethnic spatial segregation. *Urban Geography*, 19(1):77–87, 1998.

[29] David WS Wong. Geostatistics as measures of spatial segregation. *Urban geography*, 20(7):635–647, 1999.

[30] David WS Wong. Modeling local segregation: a spatial interaction approach. *Geographical and Environmental Modelling*, 6(1):81–97, 2002.