



Centrālā statistikas
pārvalde

Telpiskā statistika CSP

Dāvis Kļaviņš, ĢIS speciālists | davis.klavins@csb.gov.lv



Centrālā statistikas
pārvalde

Datu izplatīšanas telpiskā detalitāte

Publiski

- Statistiskie reģioni, administratīvās teritorijas un teritoriālās vienības
- Rīgas apkaimes (2011. gada tautas skaitīšanas (TS) dati)
- 1×1 km un 100×100 m (republikas pilsētām) režģa šūnas (galvenokārt 2011. gada TS dati)

Pētnieciskiem mērķiem

- 2000. un 2011. gada TS dati
- 2016. gada, drīzumā 2017. gada iedzīvotāju novērtējuma dati

Satur personu ID – iespējams apvienot ar administratīvajiem reģistriem



Centrālā statistikas
pārvalde

Statistiskie reģioni, administratīvās teritorijas un teritoriālās vienības

Statistiskie dati no datubāzēm:

www.csb.gov.lv/dati/statistikas-datubazes-28270.html

Telpiskie dati (ģeneralizētas robežas atbilstoši A3–A5 formātam) *shapefile* formātā:

www.csb.gov.lv/dati/kartes-un-telpiskie-dati-42749.html

Sasaiste pēc kopīgā lauka – NUTS 3 (statistiskā reģiona), LAU 2 (administratīvās teritorijas) vai teritoriālās vienības koda



Centrālā statistikas
pārvalde

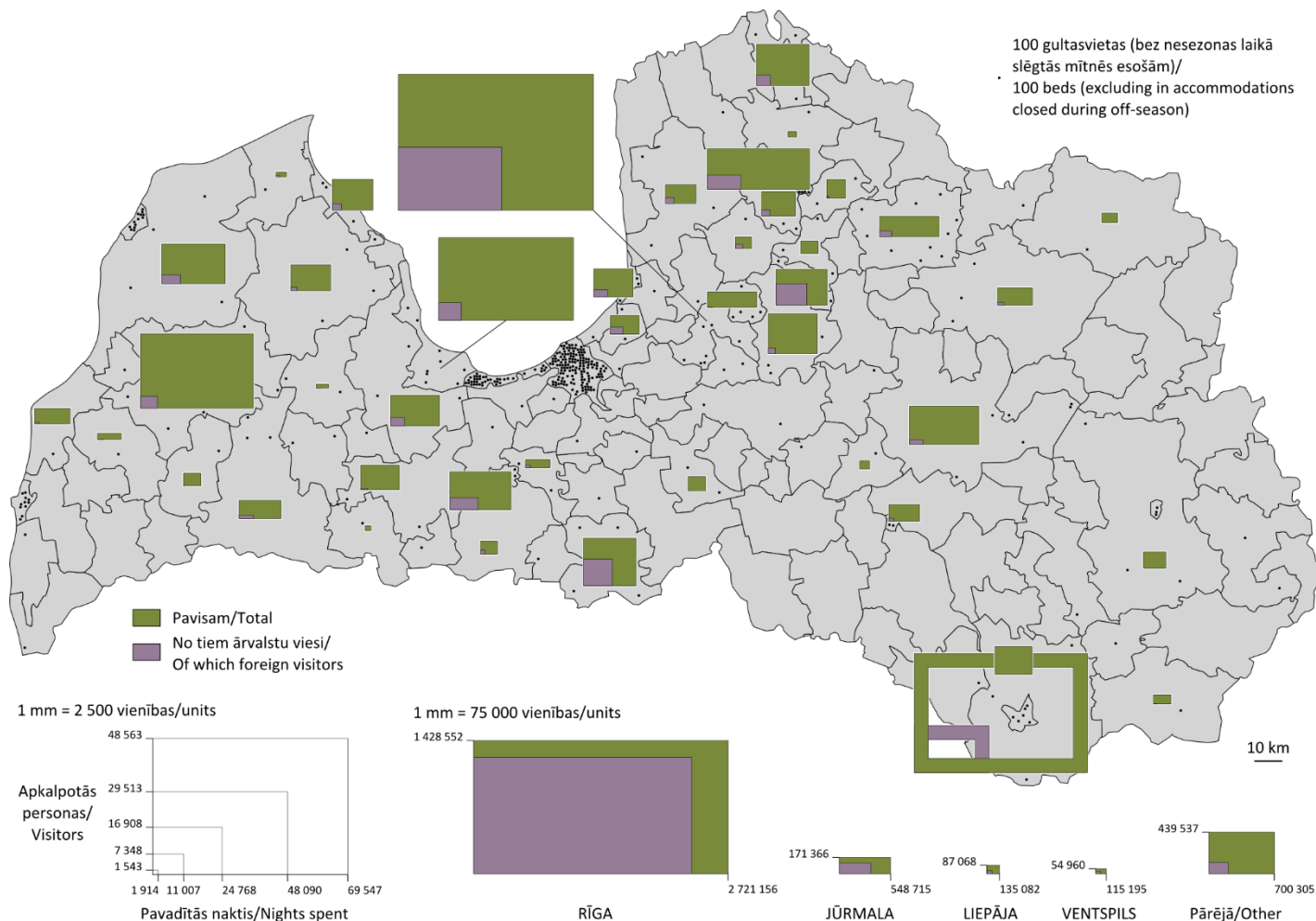
Sasaistes piemērs pēc teritoriālās vienības koda QGIS (sieviešu skaits uz 100 vīriešiem 2016. gada sākumā)

The screenshot displays the QGIS interface. On the left, the Layers Panel shows a legend for the 'ISG042_f_100_m' layer, with color-coded categories ranging from 80-86 (dark blue) to 123-139 (dark red). The main map area shows a map of Latvia with these colors applied to its administrative units. Overlaid on the map is the 'Layer Properties - 1_2400000_(A5) | Joins' dialog box. The 'Joins' tab is active, showing a table with columns: Join layer, Join field, Target field, Memory cache, Prefix, and Joined fields. The table contains one row: ISG042_f_100_m, abc.L1_code, L1_code, checked, all. Below this table is the 'Add vector join' dialog box. It has dropdown menus for 'Join layer' (ISG042_f_100_m), 'Join field' (abc.L1_code), and 'Target field' (abc.L1_code). It includes checkboxes for 'Cache join layer in virtual memory' (checked), 'Create attribute index on join field' (unchecked), and 'Choose which fields are joined' (unchecked). Under 'Choose which fields are joined', there are checkboxes for 'L1_code' and 'F_100_M'. At the bottom, there is a 'Custom field name prefix' section with an empty text box. The 'OK' and 'Cancel' buttons are at the bottom right of the dialog. The status bar at the bottom of the QGIS window shows 'Coordinate: -229088,751351', 'Scale: 1:2 400 000', 'Rotation: 0.0', and 'Render: EPSG:3059 (OTF)'.



Centrālā statistikas
pārvalde

Viesnīcas un citas tūristu mītnes administratīvajās teritorijās 2016. gadā





Centrālā statistikas
pārvalde

Rīgas apkaimes

2011. gada TS dati no datubāzes:

data.csb.gov.lv/pxweb/lv/tautassk_11/ >

2011. gada tautas skaitīšanas galīgie rezultāti – Rīga

Telpiskie dati *shapefile* formātā:

opendata.riga.lv/attistiba-un-buvnieciba.html >

Apkaimju robežu dati

Sasaiste pēc kopīgā lauka – apkaimes nosaukuma

Sagatavošanas procesā 2016. un 2017. gada
demogrāfiskie dati



Centrālā statistikas
pārvalde

1×1 km un 100×100 m (republikas pilsētām) režģa šūnas

2011. gada TS un 2016. gada iedzīvotāju
novērtējuma dati CSV formātā un telpiskie dati
(režģa pārklājums) *shapefile* formātā:

www.csb.gov.lv/dati/kartes-un-telpiskie-dati-42749.html

Sasaiste pēc kopīgā lauka – režģa šūnas
identifikatora



Centrālā statistikas
pārvalde

1×1 km un 100×100 m (republikas pilsētām) režģa šūnas

Tēmas 2011. gada TS datiem:

- Pastāvīgo iedzīvotāju skaits
- Ikdienā lietotā valoda
- Izglītība
- Ekonomiskā aktivitāte
- Mājokļi

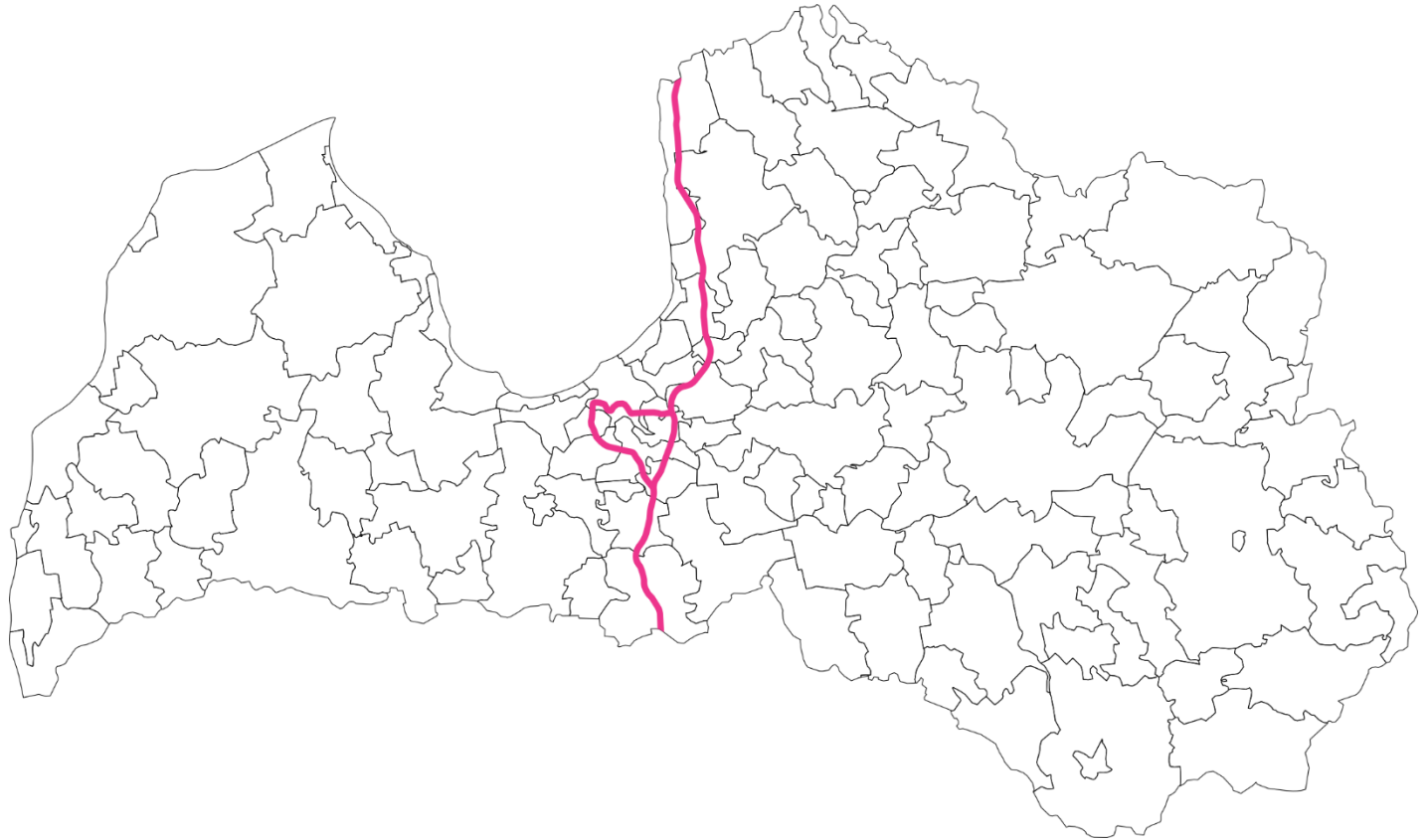
Par 2016. gadu publicēts iedzīvotāju skaits 1×1 km
režģa šūnās

Daudz plašākas telpiskās analīzes iespējas nekā
datiem par teritoriālajām vienībām



Centrālā statistikas
pārvalde

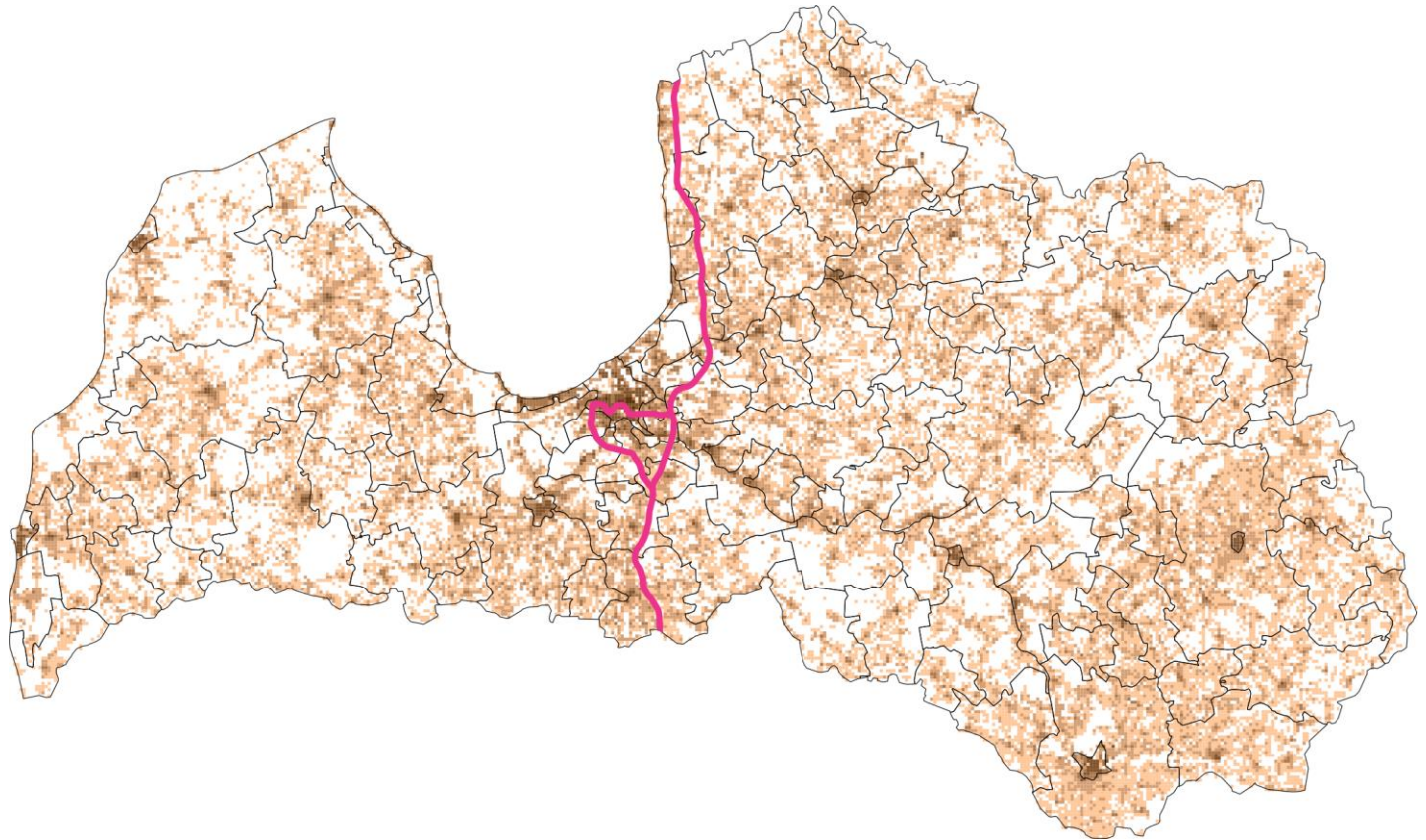
Iedzīvotāju skaita aprēķināšana izvēlētā teritorijā: 1 km ap Rail Baltica A alternatīvas sliedēm





Centrālā statistikas
pārvalde

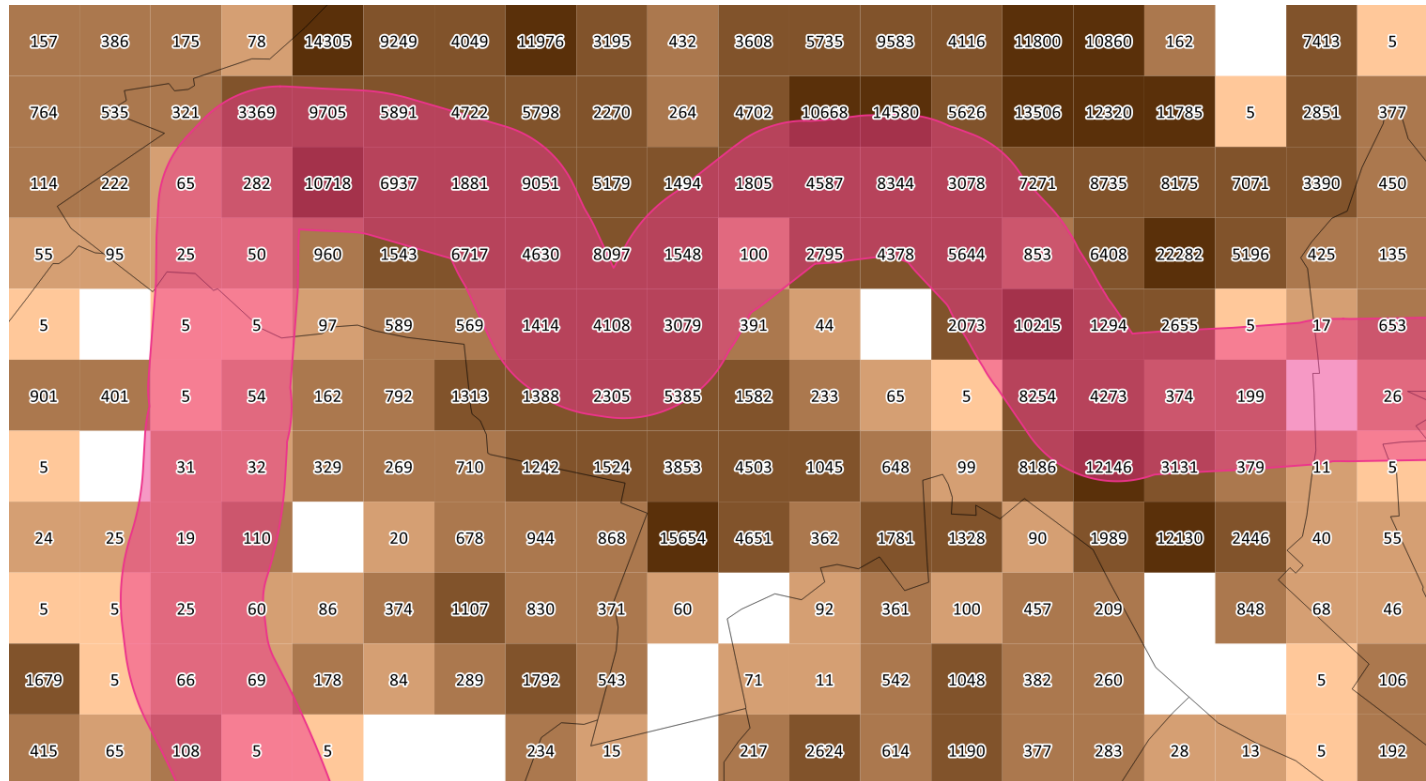
Iedzīvotāju skaita aprēķināšana izvēlētā teritorijā: 1 km ap Rail Baltica A alternatīvas sliedēm





Centrālā statistikas
pārvalde

Iedzīvotāju skaita aprēķināšana izvēlētā teritorijā: 1 km ap Rail Baltica A alternatīvas sliedēm





Centrālā statistikas
pārvalde

Iedzīvotāju skaita aprēķināšana izvēlētajā teritorijā: 1 km ap Rail Baltica A alternatīvas sliekšņiem

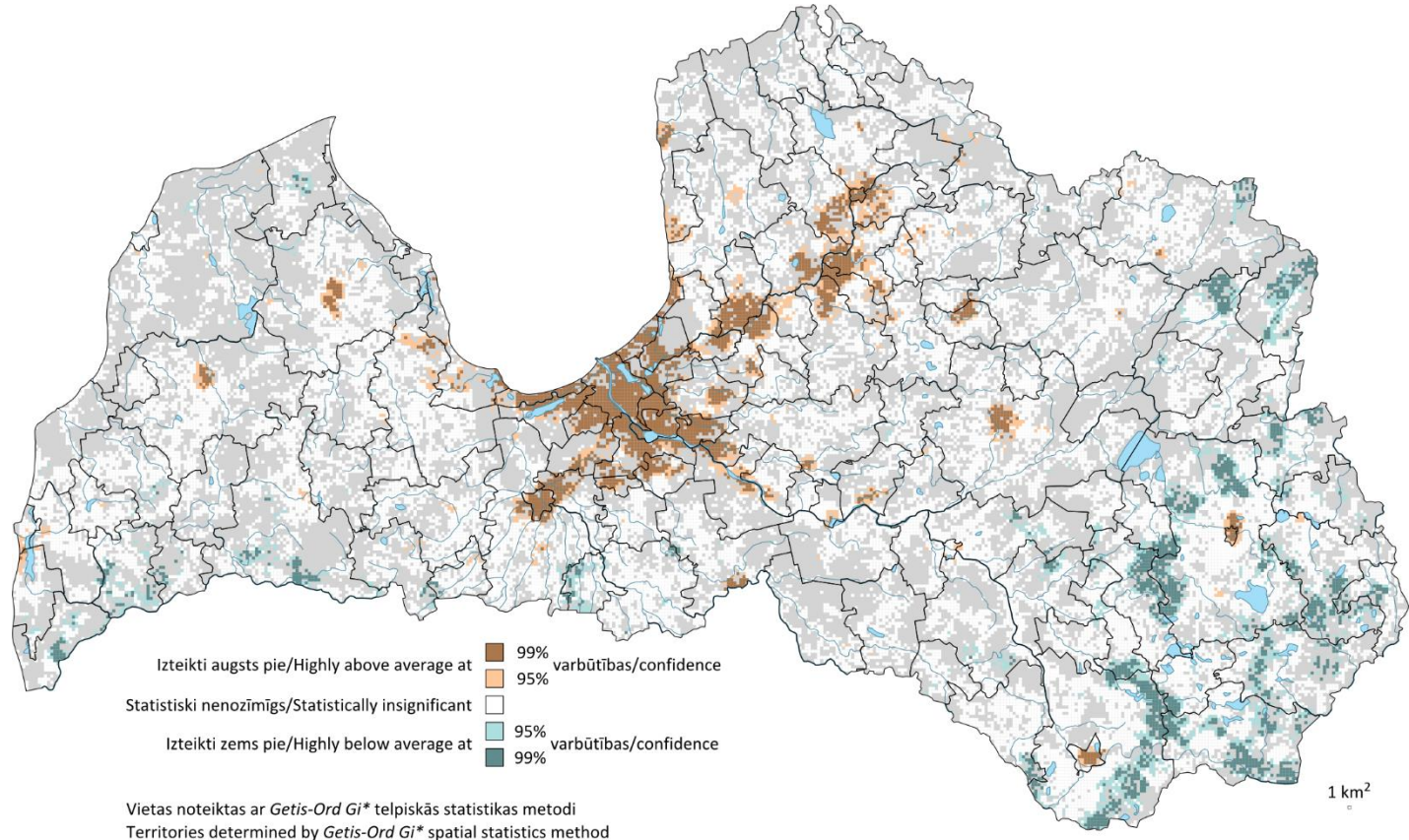
Iedzīvotāju skaits režģa šūnās, kuru lielākā daļa
atrodas izvēlētajā teritorijā – 170 279

Aprēķins veikts no publiski pieejamajiem datiem,
izmantojot individuālos datus, tas būtu precīzāks



Centrālā statistikas
pārvalde

Vietas ar izteikti augstu vai zemu vidējā izglītošanās laika indeksu vismaz 25 gadus vecu pastāvīgo iedzīvotāju vidū





Centrālā statistikas
pārvalde

Notiekošs projekts: Teritoriālās statistikas (galvenokārt pilsētas) iegūšana

Galvenie mērķi ietver:

- Statistikas nodrošināšanu republikas pilsētām, t.sk. iespēju aprēķināt datus pilsētu veidojošajām telpiskajām vienībām
- Iedzīvotāju aptauju par dzīves kvalitāti šajās pilsētās



Centrālā statistikas
pārvalde

Notiekošs projekts: Teritoriālās statistikas (galvenokārt pilsētas) iegūšana

Projekta blakusprodukts – pastāvīgo iedzīvotāju skaits 1×1 km režģa šūnās 2016. gada sākumā:
www.csb.gov.lv/statistikas-temas/kartes/pastavigo-iedzivotaju-skaitis-46019.html

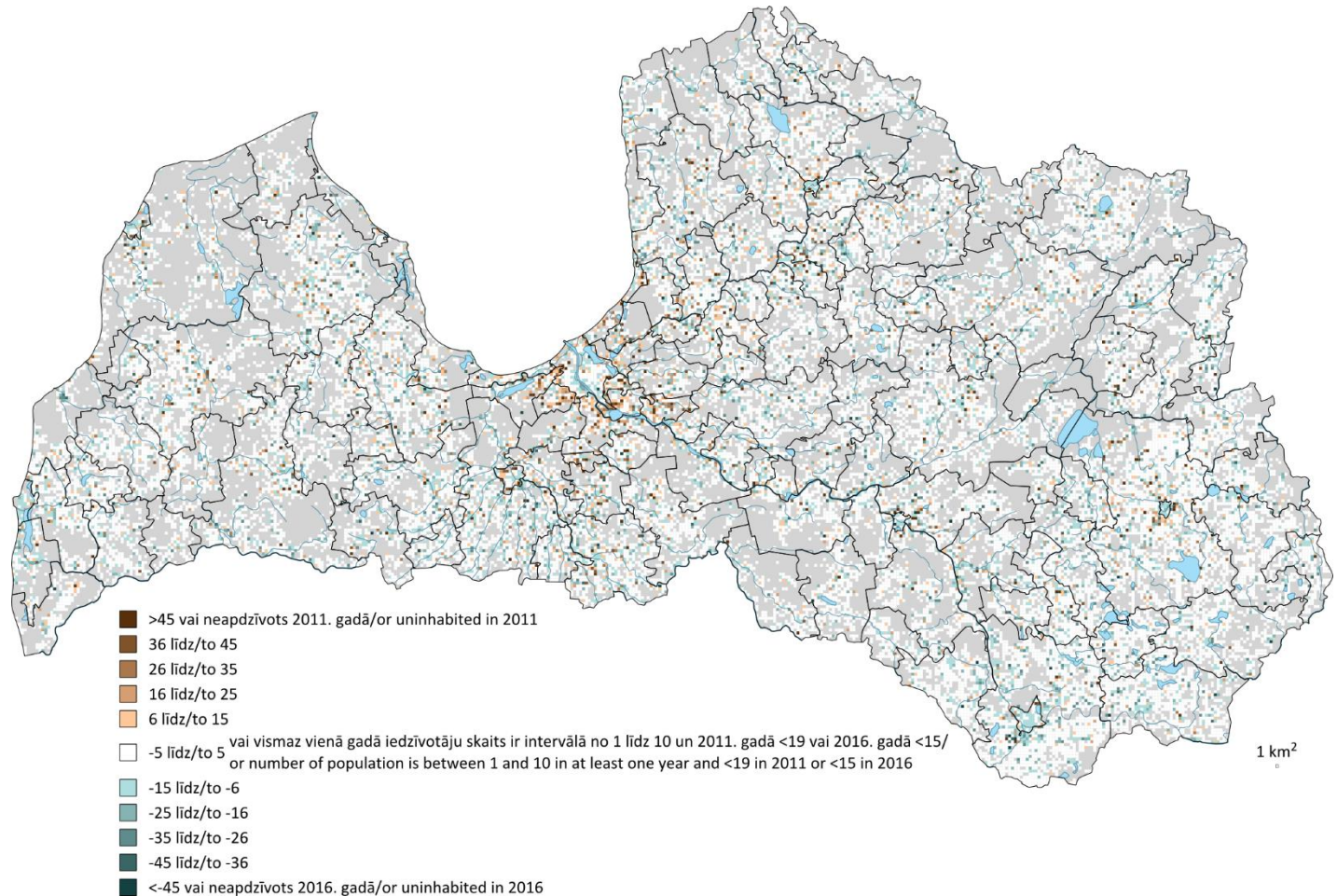
Aprēķināts, izmantojot ledzīvotāju reģistra un 2011. gada TS datus

Tiks publicēti arī dati par Rīgas apkaimēm un dati 1×1 km režģa šūnās par 2017. gada sākumu



Centrālā statistikas
pārvalde

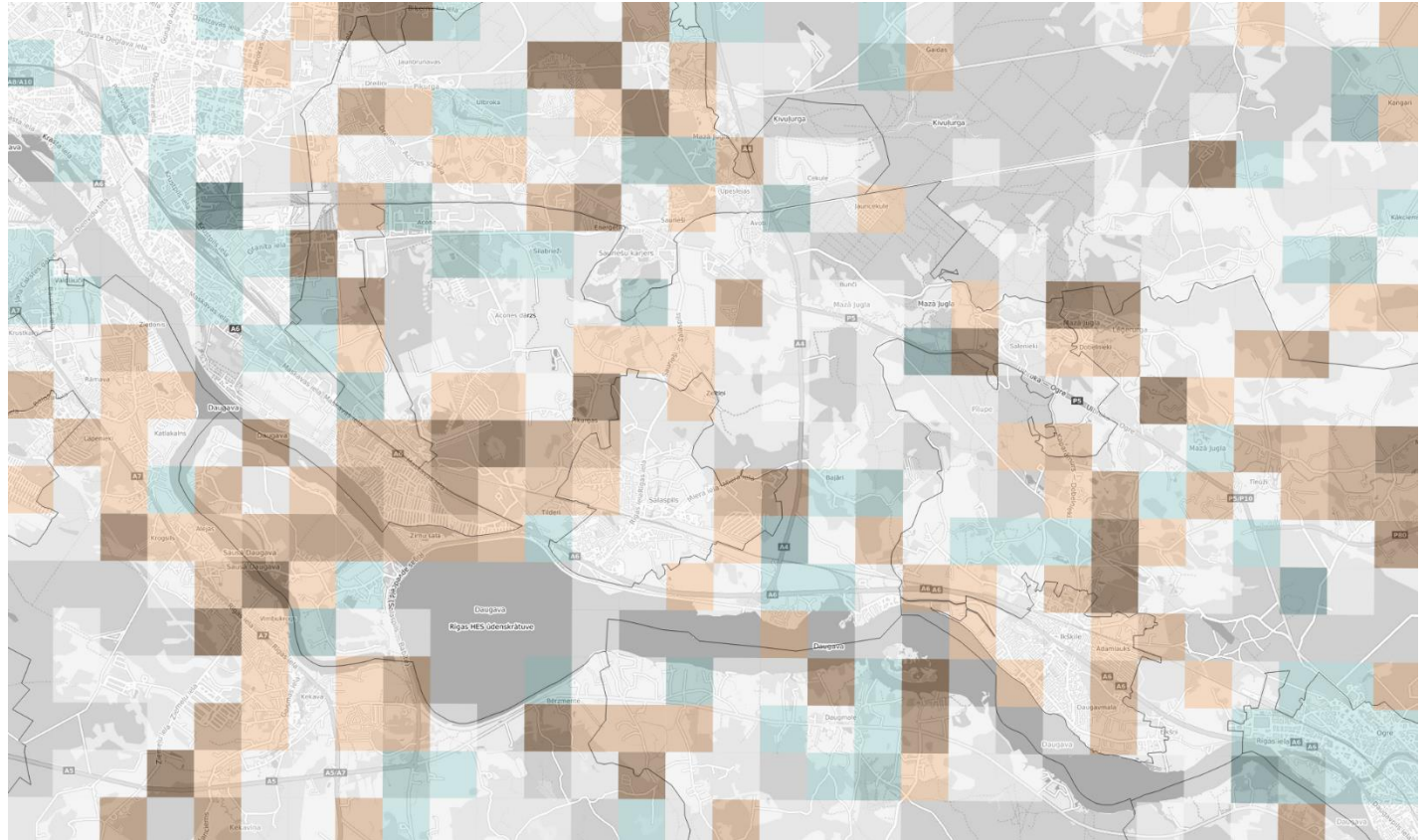
Pastāvīgo iedzīvotāju skaita izmaiņas, 2011–2016





Centrālā statistikas
pārvalde

Pastāvīgo iedzīvotāju skaita izmaiņas, 2011–2016





Centrālā statistikas
pārvalde

Notiekošs projekts: 2000. gada TS datu ģeokodēšana un migrācijas analīze

Lai veiktu telpiski detalizētus pētījumus laikā un telpā, piemēram, par migrāciju un iedzīvotāju skaita samazināšanos, nepieciešama ģeokodēta datu kopa par iedzīvotājiem, nodarbinātību un mājokļu apstākļiem. Šādus datus vispilnīgāk iespējams iegūt tautas skaitīšanās.

Atšķirībā no 2011. gada TS datiem, 2000. gada TS dati nav ģeokodēti, un to izmantošanas iespējas pētījumiem laikā un telpā ir ierobežotas



Centrālā statistikas
pārvalde

Notiekošs projekts: 2000. gada TS datu ģeokodēšana un migrācijas analīze

Projekta posmi:

- 2000. gada TS datu kopas ģeokodēšana
- Iedzīvotāju un mājojumu rādītāju publicēšana 1×1 km režģa šūnās visai Latvijas teritorijai un 100×100 m režģa šūnās republikas pilsētām, kā arī attālinātas pieejas datu kopas veidā zinātniskiem mērķiem (līdzīgi kā publicēti 2011. gada TS dati)
- Migrācijas un iedzīvotāju skaita samazināšanās analīze



Centrālā statistikas
pārvalde

Notiekošs projekts: 2000. gada TS datu ģeokodēšana un migrācijas analīze

Adrešu dati no anketām tika optiski atpazīti un satur
kļūdas:

Mājas Nr.	Ielas nosaukums
/1	TARTU
11	
11	TART4
11	TARTU
2A	TA.RTU .
2A	TAR+4
2A	TART4
2A	TARTU
2A	TART-U
-2A	TARTU
5	
5	TART4
5	TARTU
5	TARTU -
5	TARTU.
ZA	TARTU



Centrālā statistikas
pārvalde

Notiekošs projekts: 2000. gada TS datu ģeokodēšana un migrācijas analīze

Savienojot adrešu datus no 2000. gada TS un ledzīvotāju reģistra 01.03.2000., tika aprēķināta varbūtība, ka adreses sakrīt

Adrešu dati ledzīvotāju reģistrā arī satur pieraksta kļūdas, bet daudz mazāk nekā 2000. gada TS dati

Varbūtības tika aprēķinātas R programmā, izmantojot adreses pierakstu veidojošos laukus (ielas/mājas nosaukumu, mājas Nr. u.c.)



Centrālā statistikas
pārvalde

Notiekošs projekts: 2000. gada TS datu ģeokodēšana un migrācijas analīze

Katram laukam visiem ierakstiem tika aprēķināta Lēvenšteina distance (Lēvenšteina distance starp diviem vārdiem ir minimālais izmaiņu daudzums, kas nepieciešams, lai vienu vārdu pārveidotu par otru). Pēc tam tā tika izdalīta ar garākā vārda garumu un atņemta no 1, tādējādi aprēķinot varbūtību, ka lauki sakrīt.

DAUGAVPIL.5 → DAUGAVPILS (distance vienāda ar 2)

$$\text{Varbūtība} = 1 - \frac{2}{11} \approx 0.82$$



Centrālā statistikas
pārvalde

Notiekošs projekts: 2000. gada TS datu ģeokodēšana un migrācijas analīze

Varbūtības tika apvienotas ar svariem, lai aprēķinātu apvienoto varbūtību katram ierakstam

Vairākos (~30) soļos, izmantojot MS SQL Server, adreses no 2000. gada TS datu kopas tika savienotas ar 1) Valsts adrešu reģistra vēsturiskajiem datiem, 2) 2011. gada TS datu kopu (tikai adreses, kuras nesavienojās ar Valsts adrešu reģistru)

Pievienotās datu kopas satur 2000. gada TS datu kopai pievienojamās koordinātas



Centrālā statistikas
pārvalde

Notiekošs projekts: 2000. gada TS datu ģeokodēšana un migrācijas analīze

Atsevišķos soļos aprēķinātās varbūtības tika izmantotas, lai aizvietotu adresu datus no anketām ar datiem no ledzīvotāju reģistra

Atlikušajām adresēm tika aprēķināta varbūtība, ka tās sakrīt 2000. un 2011. gada TS datu kopās, un pie vairākiem nosacījumiem izmantota, lai pieņemtu, ka var tikt lietota adrese no 2011. gada TS



Centrālā statistikas
pārvalde

Notiekošs projekts: 2000. gada TS datu ģeokodēšana un migrācijas analīze

Adresēm, kuras nebija iespējams automātiski savienot un aptver ~245 000 personas (10%), tika aprēķinātas varbūtības sakritībai ar Valsts adrešu reģistra adresēm. 10 lielākās varbūtības katrai adresei tika izmantotas manuālajai savienošanai (visas, ja lielākā varbūtība sakrīt vairāk nekā 10 adresēm).

Ierakstu skaits manuālajai savienošanai ir būtiski mazāks nekā ~245 000 grupēšanas pēc adreses dēļ (sākotnēji ~37 000, bet atsevišķos gadījumos adreses nepieciešams sadalīt mājsaimniecībās vai personās)



Centrālā statistikas
pārvalde

Notiekošs projekts: 2000. gada TS datu ģeokodēšana un migrācijas analīze

Lai atvieglotu manuālo savienošanu, tika izveidota programmatūra, kurā bez varbūtībām adrešu sakritībai ar Valsts adrešu reģistra adresēm bija dota pieeja arī šādiem avotiem:

- Skanētajām anketām
- Visām intervētāja apsekotajām adresēm
- Deklarētajām dzīvesvietām līdz 2016. gadam un adresei atbilstoši 2011. gada TS



Centrālā statistikas
pārvalde

Notiekošs projekts: 2000. gada TS datu ģeokodēšana un migrācijas analīze

	Personas	% no kopējā iedzīvotāju skaita
Automātiski savienotas	2 081 263	88
Manuāli savienotas (t.sk. automātiski savienotas bez koordinātām, vairāk nekā viena automātiski pievienota adrese un izlabotas kļūdas automātiskajā savienošanā)	295 931	12



Centrālā statistikas
pārvalde

Notiekošs projekts: 2000. gada TS datu ģeokodēšana un migrācijas analīze

Ģeokodēta 2000. gada TS datu kopa palielinās tautas skaitīšanas datu analītisko potenciālu, dodot iespēju veikt pētījumus laikā un telpā, izmantojot publiski pieejamus režģa veida datus, bet zinātniskiem mērķiem arī datus personu līmenī



Centrālā statistikas
pārvalde

Režģa veida dati, būdami publiski pieejami, dod iespēju ikvienam veikt datu analīzi laikā un telpā un izmantojami gan pētniecībā un studijās, gan telpiskās attīstības plānošanā (apvienojot ar citiem datu lietotāja rīcībā esošiem datiem, tie var tikt izmantoti, piemēram, lai analizētu sabiedriskā transporta pieejamību, noteiktu vietas velosceļu izveidei vai mobilo sakaru bāzes staciju izvietojumam)



Centrālā statistikas
pārvalde

