

Visualization of Record Swapping

Jiří Novák

Prague University of Economics
and Business
Czech Statistical Office

Jaroslav Sixta

Prague University of Economics
and Business
Czech Statistical Office

Abstract

Record Swapping is a statistical disclosure control technique widely used to secure the confidentiality of microdata obtained in surveys and censuses. It is one of the methods recommended by the Centre of Excellence on Statistical Disclosure Control of the European Union for the protection of data from Census 2021. This method is based on the swapping of risk records between geographical areas. The analysis of perturbed microdata is usually done purely on numerical results, and only simplified schemes are used for general illustration. We wanted a comprehensive visualization, and therefore, we have prepared a visualization of record swapping based on Choropleth maps and Commuting Flow maps, also known as Origin-Destination maps. Choropleth maps use differing colours within predefined geographic areas to represent aggregated statistical data. Origin-Destination maps are widely used for visualization and a description of people commuting to work, but they can also be used to describe other indicators. We utilize these methods to visualize the swapping of records for individual statistical units (persons). This approach is demonstrated using microdata from the Population Census 2011 from the Czech Republic and synthetic Austrian EU-SILC data. The proposed visualization allows statistical offices and agencies to effectively evaluate the swapping process and the distribution of the records across geographical areas.

Keywords: disclosure, statistical disclosure control, SDC, record swapping, visualization, microdata, census, desire lines.

1. Introduction

Record swapping is a method of statistical disclosure control that focuses on protecting data released from national statistical offices and agencies (NSO). The basic principle of this method consists of swapping high-risk households across regions based on their risk of disclosure. Record swapping is one of the two methods which are recommended by the Centre of Excellence on Statistical Disclosure Control of the European Union by Antal, Enderle, and Giessing (2017) for the protection of the Population census 2021.

This method was initially proposed by Dalenius and Reiss (1982) and since then has been examined by researchers across the world in a significant number of articles, for example Longhurst, Tromans, Tudor, and Miller (2008), Fienberg and McIntyre (2004), Shlomo, Tudor, and Groom (2010), Soria-Comas and Domingo-Ferrer (2012), Muralidhar (2017), Lukan

and Smukavec (2017). While researching the scientific papers and studies, it is very noticeable that this method was never properly visualized; only simplified illustrations of the method are presented. The analysis of perturbed microdata is usually performed purely with numerical results, and the principle of the method has generally only been illustrated on a basic scheme picture. For example, the following illustrations can be found.

Figure 1 shows an illustrative example by the Census Division from National Records of Scotland NRS (2013) from the PAMS Conference. Figure 2 shows an illustration of targeted record swapping from Dove, Ntoumos, and Spicer (2018) where it was used to describe method used by the United Kingdom Office for National Statistics in 2011. Figure 3 shows illustrative example by Northern Ireland Statistics and Research Agency NISRA (2021) from their Census Statistical Disclosure Control Methodology.

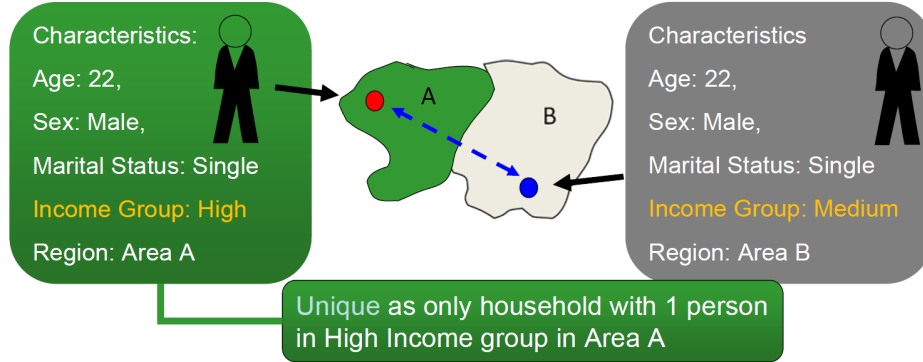


Figure 1: Illustrative example by (NRS 2013)

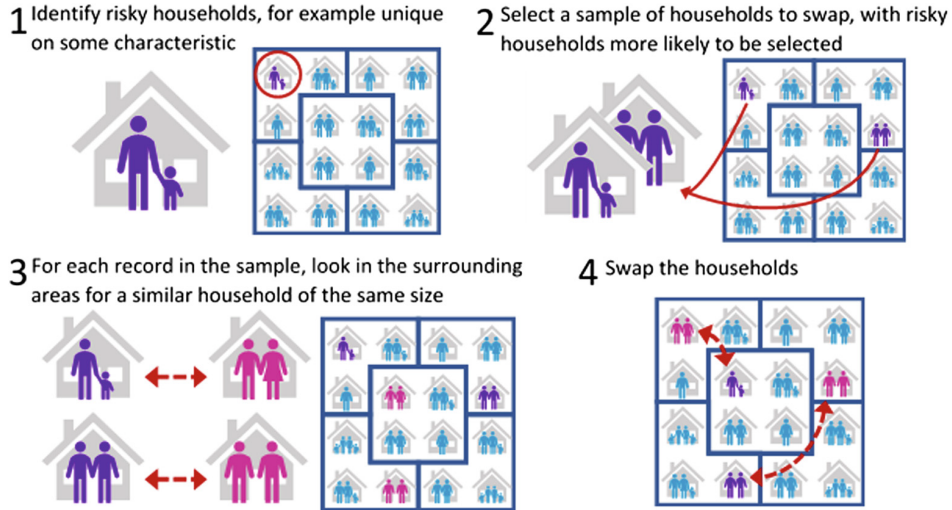


Figure 2: Illustrative example by (Dove *et al.* 2018)

We contribute to the research and scientific papers by proposing a method to enable proper visualization. In contrast to the articles listed above, where the method is only described in the text, and the results are evaluated only numerically, we were able to create a visualization through which statistical offices and agencies can observe the swapping between regions. This approach allows statistical offices and agencies to easily visually evaluate the swapping process and the distribution of the records across geographical areas.

Our paper introduces a novel approach to the visualization of record swapping. We use well-established Choropleth maps and Commuting Flow maps, also known as Origin-Destination maps. These provide an intuitive tool for understanding commuting flows, for example,

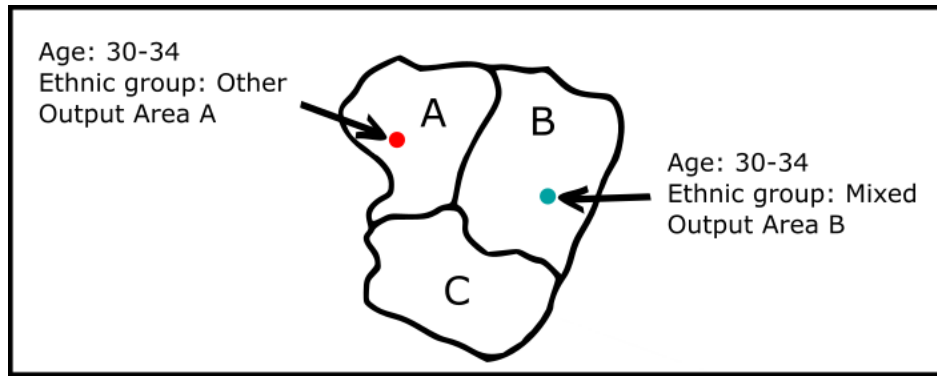


Figure 3: Illustrative example by (NISRA 2021)

commuting to work. The underlying idea of using these maps is that swapping people between regions is the same principle as if these people were commuting to those regions. If we swap a certain number of people in order to protect data, we can visualize it with maps designed to depict people's migration.

This paper is structured in three sections. The first section outlines the process of record swapping. In the second part, we have described the basics of Choropleth maps and Commuting Flow Maps. Finally, in the third section, we have illustrated the proposed visualization method using an example derived from the 2011 Population Census data from the Czech Republic and synthetic Austrian EU-SILC data.

2. Record swapping

The method of record swapping was introduced for the first time approximately forty years ago by [Dalenius and Reiss \(1982\)](#). For further description and analysis, we have considered the adjusted method recommended by the Centre of Excellence on Statistical Disclosure Control of the European Union by [Antal et al. \(2017\)](#) for the protection of the Population Census 2021, which involves targeted record swapping. The idea behind this method is quite simple. Firstly, all records comprising the risk of disclosure are detected. Subsequently, these records are swapped with similar records from other geographical areas.

For the analysis, we use R programming language by [R Core Team \(2023\)](#) with the package called `sdcMicro` by [Templ, Meindl, Kowarik, Gussenbauer, Development, Netherlands, and Heus \(2023\)](#).

As ([Shlomo et al. 2010](#)) described, this method is targeted at high-risk households, where the swap conditions are determined by the geographical hierarchies and records that are most at risk are defined by unique cells on margins of key variables ([Young, Martin, and Skinner 2009](#)). The number of swapped records is based on two factors: the proportion of unique cells defined by the combination of key variables, estimated in used analysis by the k-anonymity rule. The second factor is defined by the swap rate, a manually set value determined by the NSO. The swap rate is defined as a p% sample of the households which will always be swapped. High-risk households are calculated from identified high-risk individuals. These high-risk individuals are determined on the basis of frequency counts of unvaried distributions on a set of key variables at different geographical hierarchies.

In the R package `sdcMicro` ([Templ et al. 2023](#)) the risk associated with an individual is calculated using counts over the geographic hierarchies and the combination of all risk variables. The k-anonymity rule is used for the estimation of the risk $r_{i,h}$ for each individual i in the geographic hierarchy h and can be defined as

$$r_{i,h} = \left(\sum_{j=1}^{N_{g_1}} 1[v_{1(i)} = v_{1(j)} \wedge \dots \wedge v_{p(i)} = v_{p(j)}] \right)^{-1} ,$$

where v_1, \dots, v_p is defined as a set of risk variables, N_{g_1} being calculated as the number of persons living in the region g_1 and $1[\dots]$ as the indicator function. They describe this function more understandably in the following notation

$$r_{i,h} \sim \frac{1}{\text{counts}} .$$

Sampling probability $p_{y,h}$ for the household y at the geographic hierarchy h is calculated from the risks $r_{i,h}$ of all individuals living in the same household and is defined by the maximum risk across all household members as

$$p_{y,h} = \max_{i \text{ in household } y} (r_{i,h}) .$$

On each of the defined hierarchical levels, records of households are checked against the k-anonymity rule. If the record has a unique combination of risk variables and, therefore, does not satisfy the k-anonymity rule, the household is swapped with the so-called donor household. The households chosen as donor sets of records for given swapped households are defined by sampling probability defined above and are taken from a different geographic region at all times. This process is repeated gradually across all geographical levels from the highest hierarchy level to the lowest. The swap rate defines the total number of least swapped households from which two scenarios can arise. Firstly, there were no records left to swap that did not meet the k-anonymity rule. In this case, households are further swapped until the number of swapped records set by the swap rate is fulfilled. Secondly, more households have already been swapped than was the set swap rate, but there are still records that do not fulfil the k-anonymity rule. These records are then further swapped in order to be protected. (Templ *et al.* 2023)

For more information on targeted record swapping in census, see Antal *et al.* (2017) and Shlomo *et al.* (2010).

3. Visualization

Given the multitude of data visualization methods, we consider Choropleth maps and Origin-Destination maps the most apt techniques for effectively displaying the swapping process.

Choropleth maps (Chen, Härdle, and Unwin 2008) use differing shades or colours within pre-defined geographic areas to represent aggregated statistical data. We use classed choropleth maps that class intervals that convert continuous estimates into an ordered variable with a few values that can be represented using a few colours.

The Origin-Destination statistics are a popular output of the population census from the national statistical offices. Commuting to work or school is one of the most analyzed components of population mobility. Articles that deal with this area are, for example, Martin, Gale, Cockings, and Harfoot (2018) or Šanda (2020).

National statistical offices prepare intelligible maps for the users of the commuting statistics, also known as Origin and Destination data or just as flow data. The data (ONS 2011) represents the flow of people from one place to another and is mainly focused on Migration Statistics, Workplace Statistics, Residence Statistics and Student Statistics. Well-made interactive maps of commuting flows were prepared in Great Britain due to openly accessible

Table 1: Example of O-D data

Origin	Destination	Count
3026	3039	18696
3034	3055	12700
3042	3088	9996

dataset of [flow data from the population census 2011](#), as well as by the statistical office in Austria in their [STATatlas](#).

We have followed the procedure described by ([Lovelace, Nowosad, and Münchow 2019](#)). The Origin-Destination data can be represented by areal units, desire lines, routes, nodes or route networks. In our opinion, desire lines, which are straight lines representing records of people's travels, offer the best visualization. These lines are connected between points or zone areas in geographic space. Desire lines are created by connecting the origin's centroid and the destination zone's centroid. The name desire line comes from transport purposes as the line should represent where people desire to go between zones. The line connecting the origin and the destination zone is drawn into the map as a straight line because it should represent the quickest route between points A and B without any obstacles. Therefore, this type of visualization is best for our purposes of mapping the swapping of risk households between regions.

The data structure needed for the visualization is quite simple. The data set would be constructed with only three columns, as illustrated in table 1. The first column, "Origin", represents the swapped household's region of origin, the second column, "Destination", represents the swapped household's destination region, and the third column represents the sum of swaps from the region of origin to the destination region.

For the visualization, we have used the R programming language by [R Core Team \(2023\)](#) with the packages `tmap` by [Tennekes \(2018\)](#), `stplanr` by [Lovelace and Ellison \(2019\)](#), `tidyverse` by [Wickham, Averick, Bryan, Chang, McGowan, François, Grolemund, Hayes, Henry, Hester, Kuhn, Pedersen, Miller, Bache, Müller, Ooms, Robinson, Seidel, Spinu, Takahashi, Vaughan, Wilke, Woo, and Yutani \(2019\)](#), `RCzechia` by [Lacko \(2021\)](#). For more information on transportation geocomputation, see [Lovelace et al. \(2019\)](#). The application is further described on Population census data 2011 in the following chapter.

4. Example of the visualization

We present the proposed visualization first on the data from the Population Census 2011 from the Czech Republic. The data for displaying the visualization proposal was not randomly selected, but the idea of visualizing this method came to us while we were analysing the impacts of this method on the census data. At the end of this part, we also present the visualization of another dataset for which we have chosen synthetic Austrian EU-SILC data from the R package `laeken` by [Alfons and Templ \(2014\)](#).

For the visualization, we propose to use Choropleth maps and Commuting flow maps, also known as Origin-Destination maps. The idea behind using Origin-Destination maps is that swapping people between regions is the same principle as if these people commute for work to that region. If we exchange a certain number of people in order to protect data, we can visualize the swap with maps designed to show the migration of people.

Origin-Destination maps can be used to identify the most significant movements of households using the record swapping method. At the same time, they serve as an elegant visual check that the method behaves according to its set parameterization, for example, in the case of a condition that the household does not leave the borders of its region. Choropleth maps then show effortlessly and clearly which regions are most affected by the given method.

The following part consists of 12 maps that show the swapping of households in the Czech Republic. We ran the record swapping function of the `sdcMicro` package in the R language. The model parameters were set to the k-anonymity rule on level 3, with the swap rate set at 10 %. The geographical hierarchy parameter of the function was set at regions (NUTS 3), districts (LAU1) and municipalities (LAU 2). Risk variables, from which the k-anonymity rule was calculated, were set as variables of age category, sex, nationality, and religion. To show different situations, we ran the model two times, with the only difference between models being in the set similarity profile parameter. Both models had set variables defying the household size in similarity parameter, so the swapping between the households would occur only between households of the same size.

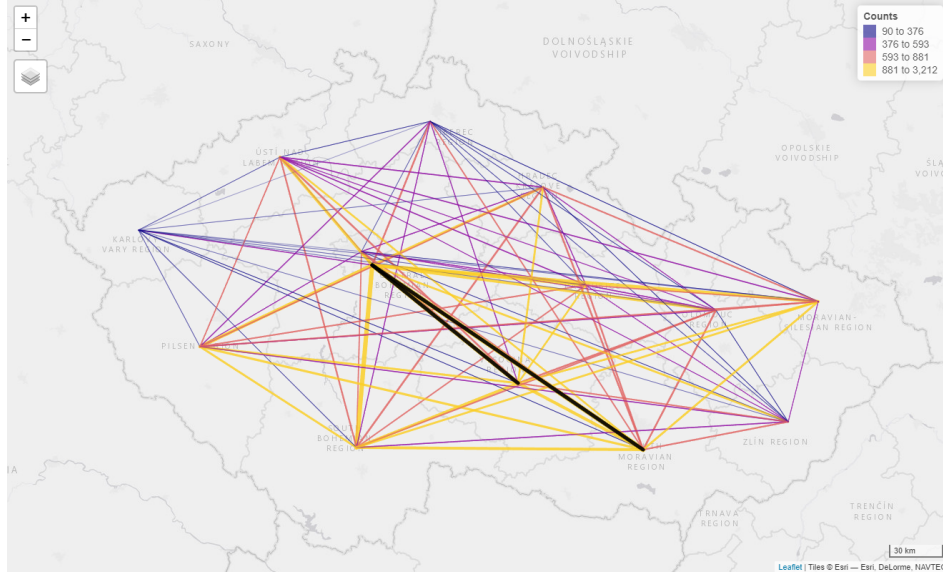


Figure 4: A map of swaps between NUTS 3 regions without the parameter of the region

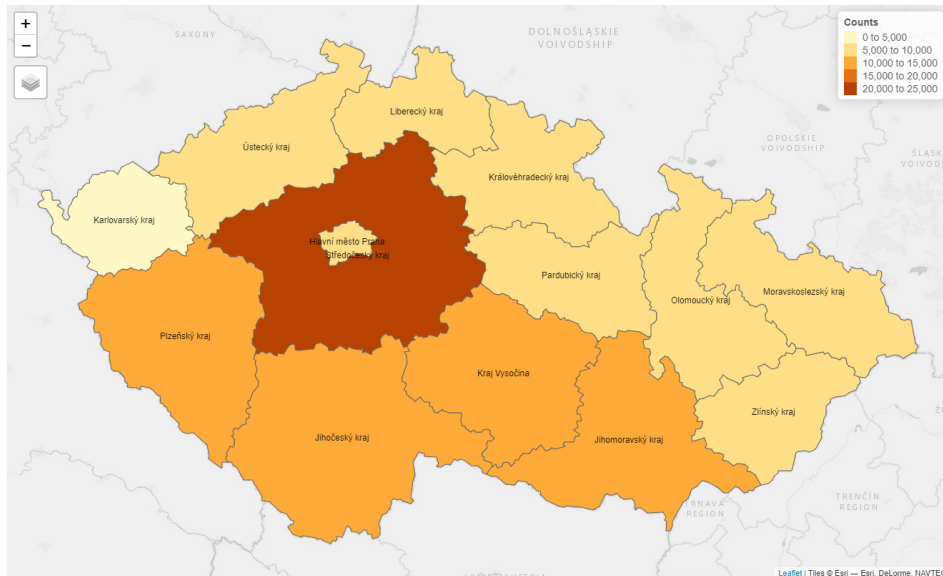


Figure 5: Choropleth map for swaps between NUTS 3 regions without the parameter of the region

Firstly, we performed the model only with this parameterization, which caused swapping households across the whole country. Secondly, the model has added the parameter of regions (NUTS 3), which stipulates that households have to stay within the borders of their respective

NUTS 3 regions. We differentiate between these two models below as models with or without the parameter of the region.

The visualization can be plotted on many different geographical hierarchy levels. We present the differences in the geographical levels of regions (NUTS 3), districts (LAU 1) and municipalities (LAU 2). For the visualization, we have used the R programming language by [R Core Team \(2023\)](#) with the packages `tmap` by [Tennekes \(2018\)](#), `stplanr` by [Lovelace and Ellison \(2019\)](#), `tidyverse` by [Wickham et al. \(2019\)](#), `RCzechia` by [Lacko \(2021\)](#), `giscoR` by [Hernangómez, EuroGeographics, and Arel-Bundock \(2023\)](#). The package `tidyverse` was used for data tidying for purposes of other functions. The package `RCzechia` provided geopolygons of the Czech Republic's geographical hierarchy. The package `giscoR` provided geopolygons of the Austrian's geographical hierarchy. The package `stplanr` has provided the main functions for the Origin-Destination analysis. The package `tmap` has been used for the final visualization of the maps.

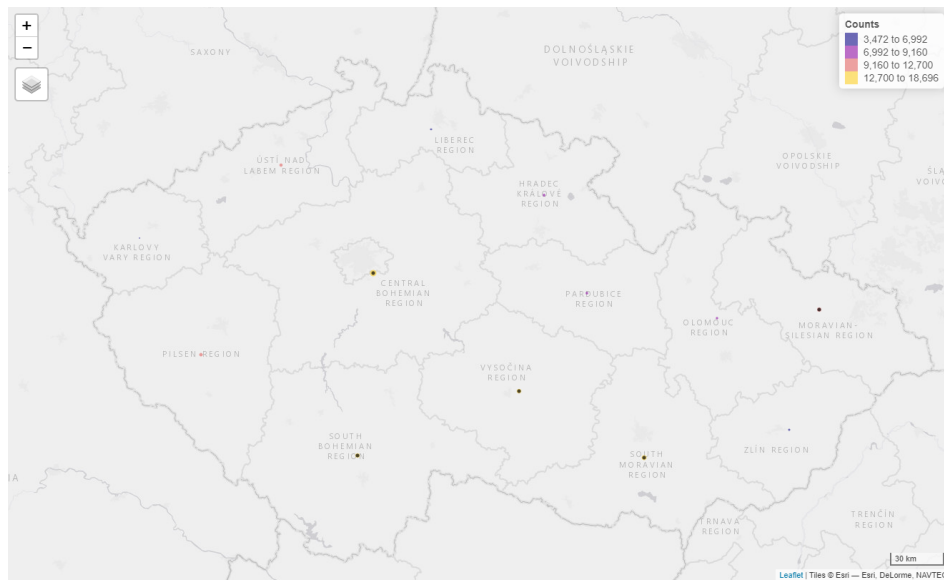


Figure 6: A map of swaps between NUTS 3 regions with the parameter of the region

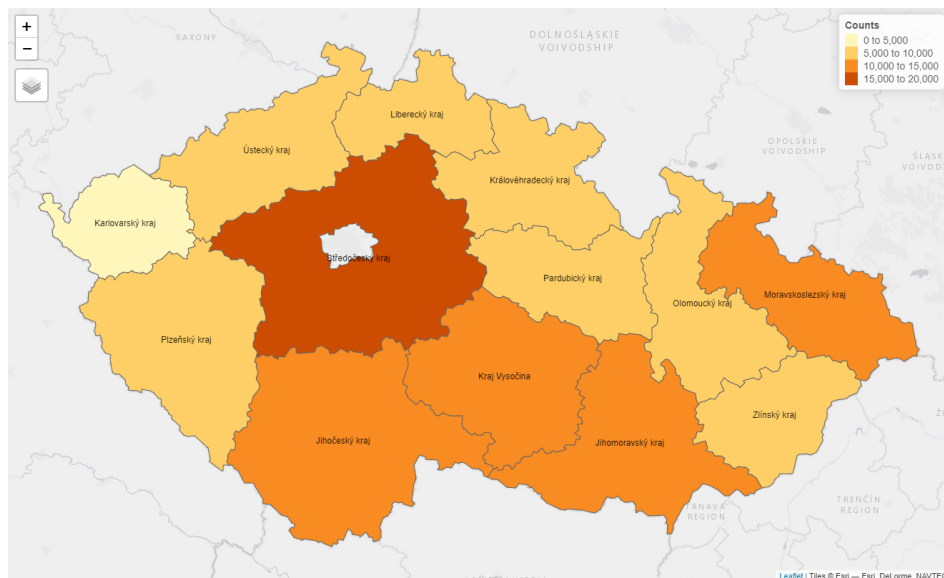


Figure 7: Choropleth map of swaps between NUTS 3 regions with the parameter of the region

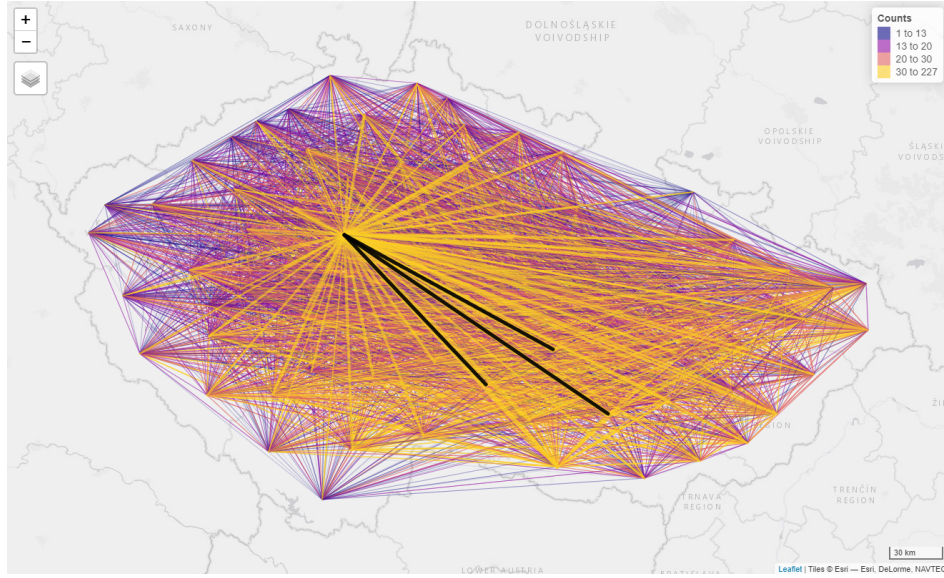


Figure 8: A map of swaps between LAU 1 districts without the parameter of the region

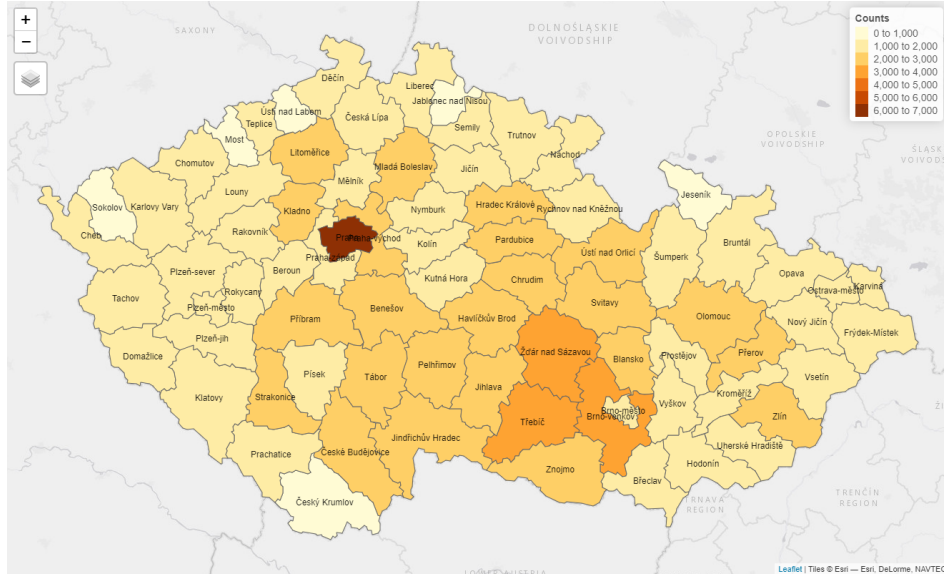


Figure 9: Choropleth map of swaps between LAU 1 districts without the parameter of the region

The visualization can be plotted on many different geographical hierarchy levels. We present the differences in the geographical levels of regions (NUTS 3), districts (LAU 1) and municipalities (LAU 2). For the visualization, we have used the R programming language by [R Core Team \(2023\)](#) with the packages `tmap` by [Tennekes \(2018\)](#), `stplanr` by [Lovelace and Ellison \(2019\)](#), `tidyverse` by [Wickham et al. \(2019\)](#), `RCzechia` by [Lacko \(2021\)](#), `giscoR` by [Hernangómez et al. \(2023\)](#). The package `tidyverse` was used for data tidying for purposes of other functions. The package `RCzechia` provided geopolygons of the Czech Republic's geographical hierarchy. The package `giscoR` provided geopolygons of the Austria's geographical hierarchy. The package `stplanr` has provided the main functions for the Origin-Destination analysis. The package `tmap` has been used for the final visualization of the maps.

Origin-Destination maps, which we here call the maps of swaps between regions, show the connections between the centroids of each region and the frequencies of swaps are colour-coded. To emphasize the swaps with the highest frequency, we have highlighted them on the



Figure 10: A map of swaps between LAU 1 districts with the parameter of the region

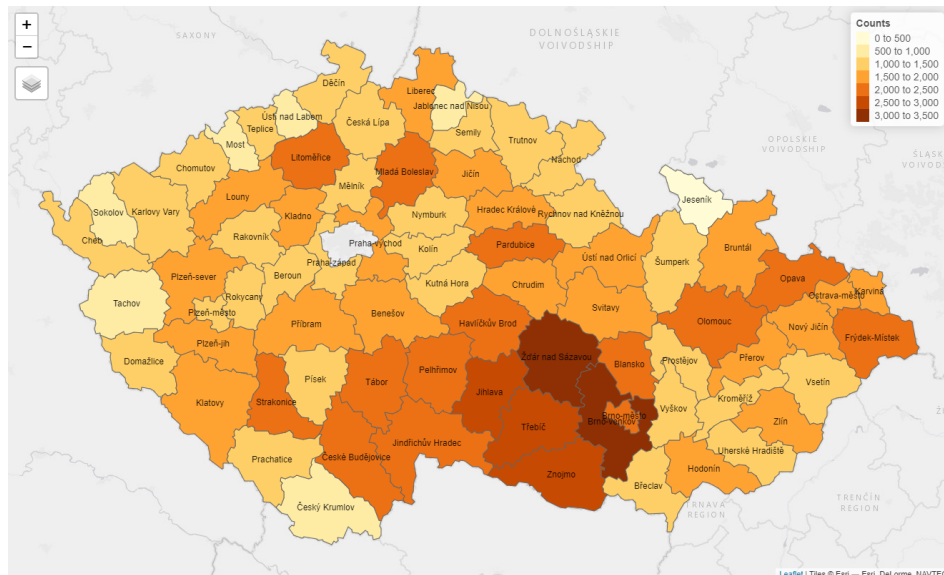


Figure 11: Choropleth map of swaps between LAU 1 districts with the parameter of the region

map with a thick black line. For the Choropleth maps, we use a simple colour scheme in shades of orange, with the intensity level set so that the lightest region represents the lowest values and the darkest region represents the most significant values for the number of swaps. These maps clearly show regions from which the highest number of people was swapped.

The data from the Population Census 2011 from the Czech Republic are used for the first part of the visualization. It is necessary to emphasize that the resulting maps do not show the final settings of the given method for the protection of census data in the Czech Republic. The given maps are only illustrative in the hypothetical use of the given method.

Figure 4 shows a map of swaps between NUTS 3 regions without applying the condition that the household must stay within the borders of the NUTS 3 region. We can see that most households are swapped between the Stredocesky region with the Jihomoravsky region and the Stredocesky region with the Vysocina region. The most swapped households are from the Stredocesky region. The lowest swapping rate can be observed in the Karlovarsky region.

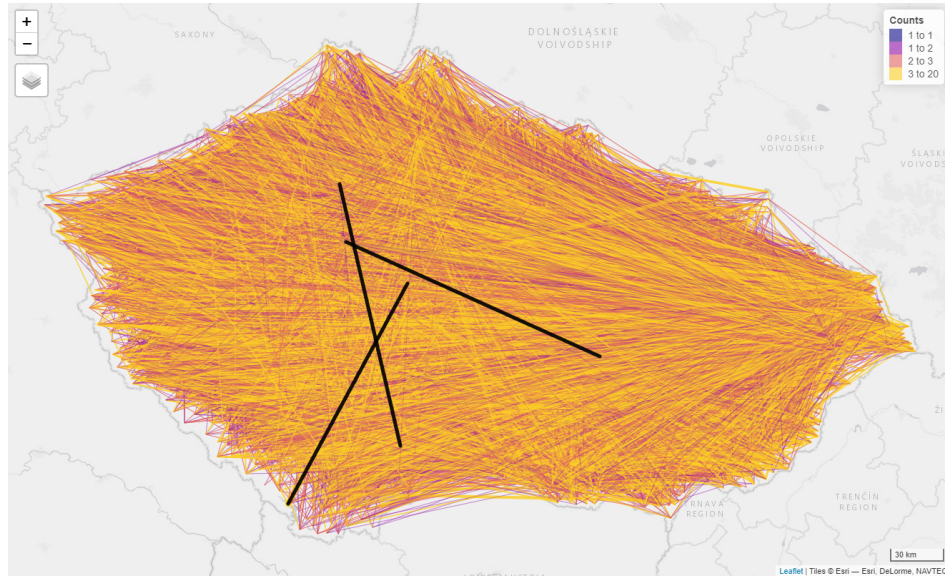


Figure 12: A map of swaps between LAU 2 municipalities without the parameter of the region

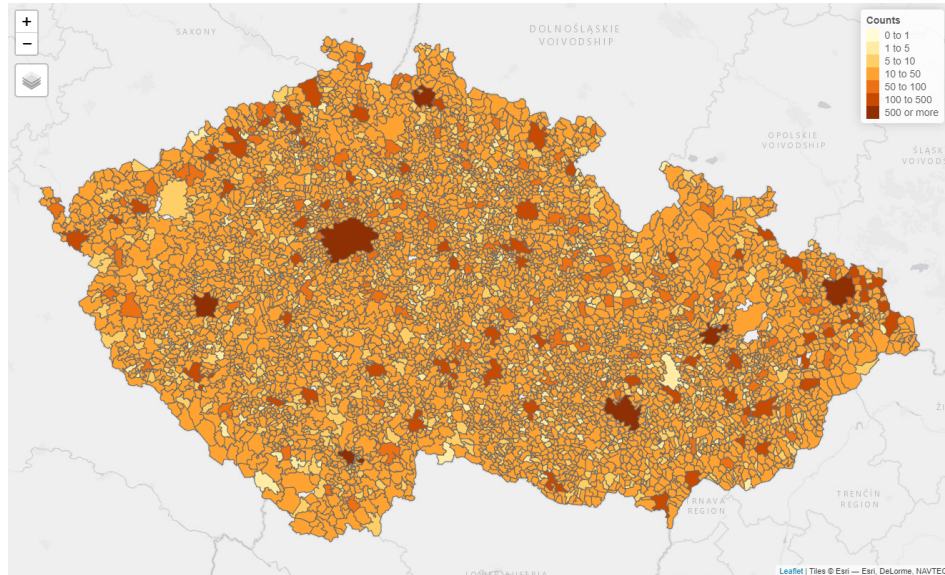


Figure 13: Choropleth map of swaps between LAU 2 municipalities without the parameter of the region

Figure 5 shows the choropleth map for counts of swap between NUTS 3 regions without application of the given condition. The most swapped households were in the Stredocesky region and the least in the Karlovarsky region.

In contrast, there is figure 4 that shows a map of swaps between NUTS 3 regions with the condition. If we apply the condition that households have to remain within the borders of the NUTS 3 region while we plot the swapping lines at the NUTS 3 region, we shall correctly observe an empty map because otherwise, the condition would be violated. More information can be seen in figure 7, which shows the choropleth map for counts of swaps between NUTS 3 regions with the application of the given condition. Prague region is missing from the map because Prague is its own region; therefore, the swapping did not take place. The most swapped households were again in the Stredocesky region and the least in the Karlovarsky region. Compared with Figure 5, the Moravskoslezsky region had more swaps and Plzensky region had fewer.

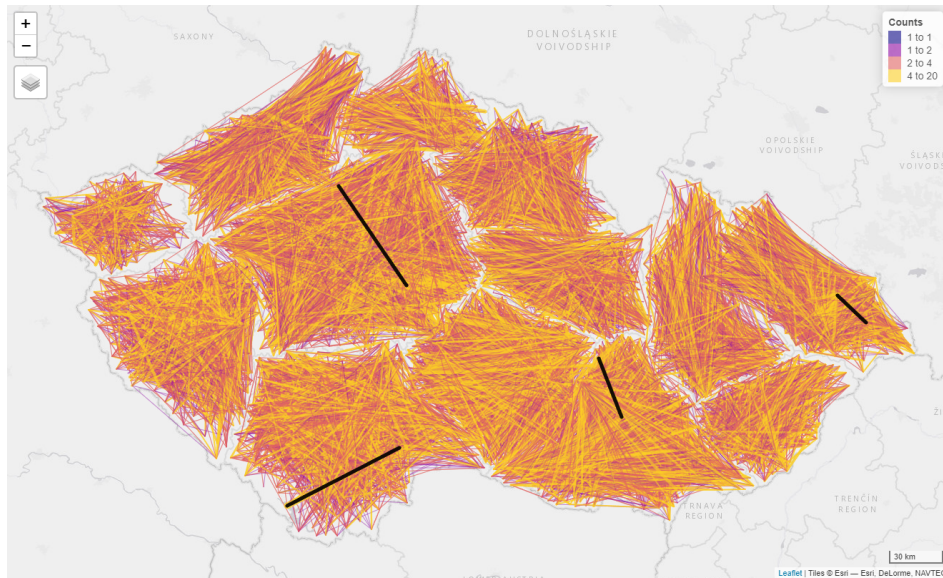


Figure 14: A map of swaps between LAU 2 municipalities with the parameter of the region

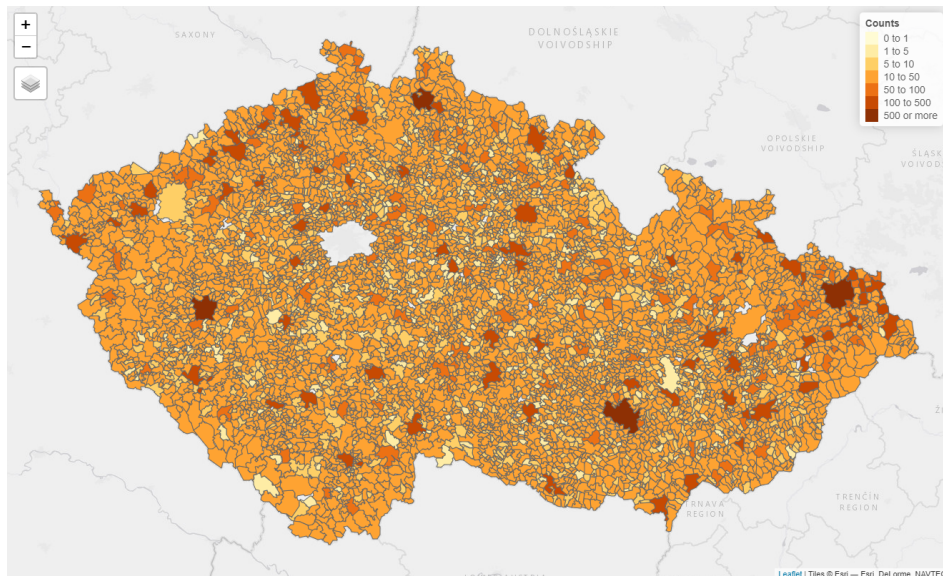


Figure 15: Choropleth map of swaps between LAU 2 municipalities with the parameter of the region

If we move to the lower level of LAU 1 districts, we can analyse figure 8, which shows a map of swaps between LAU 1 districts without applying the parameter of the region. On this level, clarity begins to fade with growing details. However, we observe from the map that the largest swaps occur between the centre of the country and its eastern parts. In contrast, western districts have darker lines, indicating lower swapping rates. The capital is apparent on the map with the highest swapping lines. In the choropleth map in figure 9, we can see that the most significant number of swaps occurred in the biggest city of the Czech Republic, Prague, but the second largest city, Brno (district Brno-mesto), is at the same level as the rest of the districts. Interestingly, there were more swaps around Brno (district Brno-venkov) and near Brno (district Trebic and Zdar nad Sazavou).

In contrast, we have figure 10, which shows a map of swaps between LAU 1 districts with the parameter of the region. The effect of the capital is now clearly eliminated because, in the Czech Republic, the capital is its own region. The biggest swaps can be observed in the

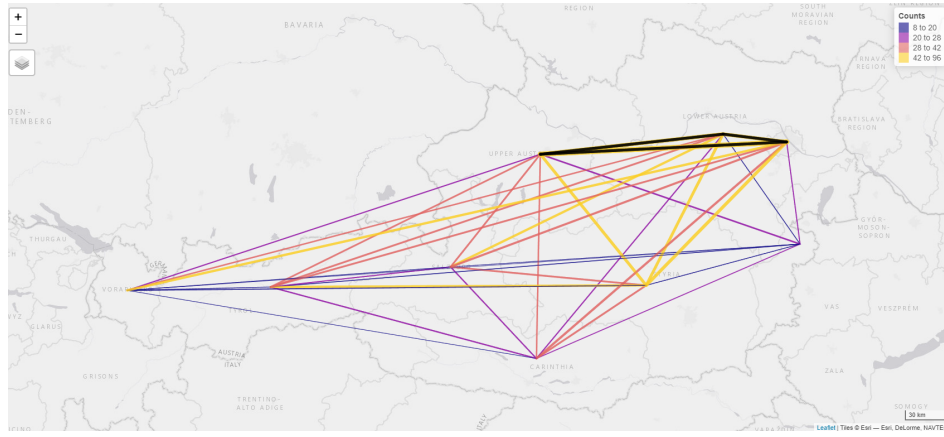


Figure 16: A map of swaps between NUTS 2 regions in Austria

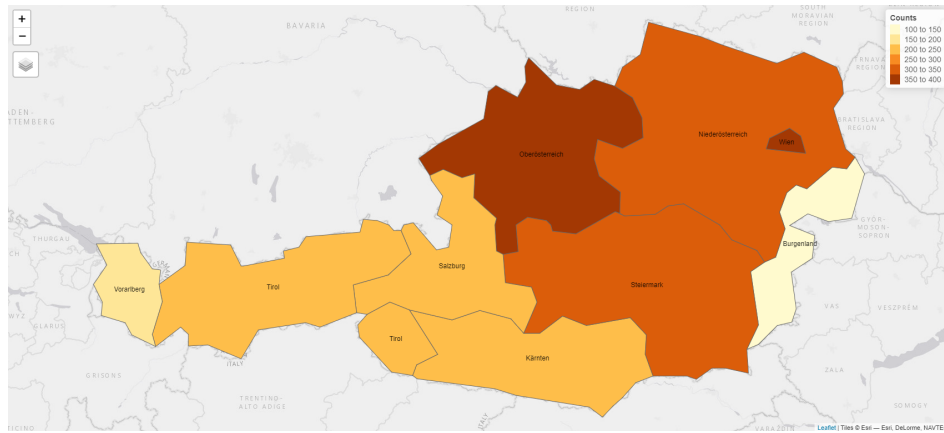


Figure 17: Choropleth map of swaps between NUTS 2 regions of Austria

middle of the country, with the highest in the Vysocina region. The lowest swap rates are in the district of the Stredocesky region. In the choropleth map figure 11, in the Vysocina region, most people were swapped between district Zdar nad Sazavou and district Trebic. Also, as in figure 9, most people were swapped in the Jihocesky region in the districts around Brno. Interestingly, districts in northern regions show fewer swaps than districts in southern regions.

The lowest level of LAU 2 municipalities is exceptionally cluttered. This is the situation of figure 12, which shows a map of swaps between LAU 2 municipalities without the condition. Although the highest swaps are located in the middle of the republic, there is a clear pattern of swaps between the west and east parts of the republic. This would be unacceptable if applied to the final data for dissemination. At the level of municipalities in figure 13, the largest cities with the largest number of swaps are coloured the most.

In contrast, the last map, figure 14, shows a map of swaps between LAU 2 municipalities with the condition. Here, it can be seen that in each region, based on the colour perception of the swap lines, there is approximately the same level of swapping. It is not evident that any region would be overestimated or underestimated. The largest swapping occurs in the areas with the largest populations, so no suspicious movements are indicated. At the level of municipalities in figure 15, the largest cities where the most swaps took place are coloured the most. Interestingly, in the Moravian-Silesian region, there are quite a number of municipalities with a high number of swaps.

The last maps use synthetic Austrian EU-SILC data. We present the swapping process at the

geographical level of NUTS 2, which are federal states. As risk variables, three variables were used: a person's age, economic status, and citizenship. The swap rate was set to 5 %, and only households of the same size were swapped. As figure 16 and figure 17 show, most people were swapped from Viena to its close regions, Upper Austria, Lower Austria and Styria. The fewest persons were swapped in the Vorarlberg and Burgenland regions.

5. Conclusion

We propose a new approach for the visualization of record swapping. Targeted record swapping is a method recommended by the EU Centre for Excellence in Statistical Confidentiality Protection for microdata protection for the 2021 Census of Population, Housing and Dwellings. However, in scientific papers, the analysis of perturbed microdata using the record swapping method is always performed purely on numerical results. It means only comparing the difference between the original microdata and the new protected microdata.

For the visualization, we have used the Choropleth maps and Commuting Flow maps, also known as the Origin-Destination maps. These maps serve as an intuitive tool for understanding commuting flows, for example, commuting to work. The idea behind these maps is that we can use the same principle for swapping people between regions as if these people commuted to that region. If we exchange a certain number of people in order to protect microdata, we can visualize such swaps with maps designed to show the migration of people.

We find this depiction of the swapping of people very beneficial. Thanks to simple visualization of the Origin-Destination lines, we can track suspicious swap movements and unwanted patterns that may occur during swapping based on a visual evaluation of the map. We consider greater swapping of households from one corner of the republic to another to be highly undesirable, and thanks to the proposed visualization, we can quickly trace this undesirable behaviour. These maps have proven to be very useful in communication with statisticians when enforcing the highest level of statistical disclosure control. Of course, the outcomes of the record swapping can be presented to the management for a more straightforward decision-making process.

The result of the proposed visualization design is a map of the country where the microdata protection process is plotted. This will simplify the process for employees of national statistical offices to visually evaluate the protection process, enhancing the security of disseminated data and increasing its availability for further research. Visualisation of record swapping will provide both an explanation of perturbed microdata and the quality control of this process.

6. Acknowledgment

This paper has been prepared with the support of a project of the Prague University of Economics and Business – Internal Grant Agency, project No. F4/50/2021.

References

- Alfons A, Templ M (2014). “Estimation of Social Exclusion Indicators from Complex Surveys. The R Package laeken.” *Journal of Statistical Software*, p. 54. URL <https://rdr.io/cran/laeken/f/inst/doc/laeken-intro.pdf>.
- Antal L, Enderle T, Giessing S (2017). *Harmonised Protection of Census Data in the ESS*. URL https://ec.europa.eu/eurostat/cros/system/files/methods_for_protecting_census_data.pdf.
- Chen Ch, Härdle W, Unwin A (eds.) (2008). *Handbook of Data Visualization*. Springer

- Handbooks of Computational Statistics. Springer, Berlin. ISBN 978-3-540-33036-3. OCLC: ocm76799021.
- Dalenius T, Reiss SP (1982). “Data-swapping: A Technique for Disclosure Control.” *Journal of Statistical Planning and Inference*, **6**(1), 73–85. ISSN 03783758. doi: [10.1016/0378-3758\(82\)90058-1](https://doi.org/10.1016/0378-3758(82)90058-1). URL <https://linkinghub.elsevier.com/retrieve/pii/0378375882900581>.
- Dove I, Ntoumos C, Spicer K (2018). “Protecting Census 2021 Origin-destination Data Using a Combination of Cell-key Perturbation and Suppression.” In J Domingo-Ferrer, F Montes (eds.), *Privacy in Statistical Databases*, volume 11126, pp. 43–55. Springer International Publishing, Cham. ISBN 978-3-319-99770-4 978-3-319-99771-1. doi: [10.1007/978-3-319-99771-1_4](https://doi.org/10.1007/978-3-319-99771-1_4). Series Title: Lecture Notes in Computer Science, URL http://link.springer.com/10.1007/978-3-319-99771-1_4.
- Fienberg SE, McIntyre J (2004). “Data Swapping: Variations on a Theme by Dalenius and Reiss.” In T Kanade, J Kittler, JM Kleinberg, F Mattern, JC Mitchell, M Naor, O Nierstrasz, C Pandu Rangan, B Steffen, M Sudan, D Terzopoulos, D Tygar, MY Vardi, G Weikum, J Domingo-Ferrer, V Torra (eds.), *Privacy in Statistical Databases*, volume 3050, pp. 14–29. Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 978-3-540-22118-0 978-3-540-25955-8. doi: [10.1007/978-3-540-25955-8_2](https://doi.org/10.1007/978-3-540-25955-8_2). Series Title: Lecture Notes in Computer Science, URL http://link.springer.com/10.1007/978-3-540-25955-8_2.
- Hernangómez D, EuroGeographics, Arel-Bundock V (2023). “giscoR: Download Map Data from GISCO API - Eurostat.” URL <https://cran.r-project.org/package=giscoR>.
- Lacko J (2021). *rczechia: Spatial Objects of the Czech Republic*. R package version 1.8.3, URL <https://CRAN.R-project.org/package=RCzechia>.
- Longhurst J, Tromans N, Tudor C, Miller C (2008). “Statistical Disclosure Control for the 2011 UK Census.” URL https://www.researchgate.net/publication/250865830_Statistical_Disclosure_Control_for_the_2011_UK_Census.
- Lovelace R, Ellison R (2019). “stplanr: A Package for Transport Planning.” *The R Journal*, **10**(2), 7. ISSN 2073-4859. doi: [10.32614/RJ-2018-053](https://doi.org/10.32614/RJ-2018-053). URL <https://journal.r-project.org/archive/2018/RJ-2018-053/index.html>.
- Lovelace R, Nowosad J, Münchow J (2019). *Geocomputation with R*. CRC Press, Taylor and Francis Group, CRC Press is an imprint of the Taylor and Francis Group, an informa Buisness, A Chapman & Hall Book, Boca Raton. ISBN 978-1-138-30451-2.
- Lukan J, Smukavec A (2017). “Statistical Disclosure Control Considerations of Publishing Data on Grid Squares and Territorial Units in the Application Stage.” URL https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2017/4_grid_squares.pdf.
- Martin D, Gale C, Cockings S, Harfoot A (2018). “Origin-destination Geodemographics for Analysis of Travel to Work Flows.” *Computers, Environment and Urban Systems*, **67**, 68–79. ISSN 01989715. doi: [10.1016/j.compenvurbsys.2017.09.002](https://doi.org/10.1016/j.compenvurbsys.2017.09.002). URL <https://linkinghub.elsevier.com/retrieve/pii/S0198971516303301>.
- Muralidhar K (2017). “Record Re-identification of Swapped Numerical Microdata.” *Journal of Information Privacy and Security*, **13**(1), 34–45. ISSN 1553-6548, 2333-696X. doi: [10.1080/15536548.2017.1281602](https://doi.org/10.1080/15536548.2017.1281602). URL <https://www.tandfonline.com/doi/full/10.1080/15536548.2017.1281602>.
- NISRA (2021). *2021 Census Statistical Disclosure Control Methodology*. URL <https://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/statistical-disclosure-control-methodology-for-2021-census.pdf>.

- NRS (2013). “Census Statistical Disclosure Control - the Use of Record-swapping to Protect Data Confidentiality.” URL <https://www.nrscotland.gov.uk/files/statistics/seminars/Scotlands-People-28-November/12-pams-conf-nov2013-census-disc-control.pdf>.
- ONS (2011). *2011 Census Origin-destination Data User Guide*. URL <http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-prospectus/release-plans-for-2011-census-statistics/subsequent-releases-of-specialist-products/flow-data/origin-destination-data--user-guide.pdf>.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Shlomo N, Tudor C, Groom P (2010). “Data Swapping for Protecting Census Tables.” In D Hutchison, T Kanade, J Kittler, JM Kleinberg, F Mattern, JC Mitchell, M Naor, O Nierstrasz, C Pandu Rangan, B Steffen, M Sudan, D Terzopoulos, D Tygar, MY Vardi, G Weikum, J Domingo-Ferrer, E Magkos (eds.), *Privacy in Statistical Databases*, volume 6344, pp. 41–51. Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 978-3-642-15837-7 978-3-642-15838-4. doi:10.1007/978-3-642-15838-4_4. Series Title: Lecture Notes in Computer Science, URL http://link.springer.com/10.1007/978-3-642-15838-4_4.
- Soria-Comas J, Domingo-Ferrer J (2012). “Probabilistic k-anonymity through Microaggregation and Data Swapping.” In *2012 IEEE International Conference on Fuzzy Systems*, pp. 1–8. IEEE, Brisbane, Australia. ISBN 978-1-4673-1506-7 978-1-4673-1507-4 978-1-4673-1505-0. doi:10.1109/FUZZ-IEEE.2012.6251280. URL <http://ieeexplore.ieee.org/document/6251280/>.
- Templ M, Meindl B, Kowarik A, Gussenbauer J, Development OFECO, Netherlands S, Heus P (2023). “sdcMicro: Statistical Disclosure Control Methods for Anonymization of Data and Risk Estimation.” URL <https://CRAN.R-project.org/package=sdcMicro>.
- Tennekes M (2018). “**tmap** : Thematic Maps in R.” *Journal of Statistical Software*, **84**(6). ISSN 1548-7660. doi:10.18637/jss.v084.i06. URL <http://www.jstatsoft.org/v84/i06/>.
- Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, **4**(43), 1686. ISSN 2475-9066. doi:10.21105/joss.01686. URL <https://joss.theoj.org/papers/10.21105/joss.01686>.
- Young C, Martin D, Skinner C (2009). “Geographically Intelligent Disclosure Control for Flexible Aggregation of Census Data.” *International Journal of Geographical Information Science*, **23**(4), 457–482. ISSN 1365-8816, 1362-3087. doi:10.1080/13658810801949835. URL <http://www.tandfonline.com/doi/abs/10.1080/13658810801949835>.
- Šanda R (2020). “The Rate of Response for the Topic of Commuting in the 2011 Population and Housing Census and the Reconstruction of the Response Rate Using the Jaro-Winkler Method.” *Demografie, Review for Population Research*, **2020**(62), 27–42. ISSN 1805-2991.

Affiliation:

Jiří Novák

Prague University of Economics and Business

nám. Winstona Churchilla 1938/4, 130 67 Prague

Czech Republic

E-mail: xnovj159@vse.cz

URL: <https://kest.vse.cz/ing-jiri-novak/>