CONFERENCE OF EUROPEAN STATISTICIANS
Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality Session 4
29-31 October 2019, the Hague, the Netherlands 24 September 2019

The Potential of Anonymization Methods for Creating Detailed Geographical Data in Japan

Shinsuke Ito (Chuo University, Japan), Masayuki Terada (NTT DOCOMO, INC, Japan)
ssitoh@tamacc.chuo-u.ac.jp, teradam@nttdocomo.com

Abstract and Paper

Several empirical studies on the effectiveness of disclosure limitation methods such as microaggregation, additive noise, and data swapping for official microdata were conducted by Ito and Murata (2011), Ito and Hoshino (2012, 2013, 2014). Empirical research on using top coding and recoding to create anonymized microdata that contain more detailed geographical information was conducted by Ito et al. (2015, 2016). Empirical research on adaptability of perturbative methods such as data swapping and PRAM (=Post RAndomization Methods) used to reduce disclosure risk for official microdata was conducted by Ito et al. (2017, 2018). As part of preparations for the 2020 U.S. Population Census, the U.S. Census Bureau has developed a new methodology of 'differential privacy' for the release of official statistical tables. This method creates privacy-preserved official microdata based on the concept of differential privacy. The adaptability of differential privacy to official statistical data in Japan is expected to become a topic in the future. In this paper, we survey the current discussion about differential privacy used for creating and releasing official statistical data, including the discussion in the United States. We also investigate the possibility of adapting differential privacy for detailed geographical data from the Japanese Population Census, and examine the potential of differential privacy as an anonymization method for Japanese statistical data. In an era of big data and artificial intelligence, the risk of sophisticated disclosure attacks is increasing. If differential privacy can be applied to detailed geographical data for official statistics in Japan, it can potentially strengthen Japanese statistics against such attacks.

Potential of Anonymization Methods for Creating Detailed Geographical Data in Japan

Shinsuke Ito* and Masayuki Terada**

* Faculty of Economics, Chuo University, 742-1 Higashinakano, Hachioji, Tokyo, 192-0393 Japan, E-mail: ssitoh@tamacc.chuo-u.ac.jp
** Research Laboratories, NTT DOCOMO, Inc, 3-6 Hikarino-oka, Yokosuka, Kanagawa, 239-8536 Japan E-mail: teradam@nttdocomo.com

Abstract: Several empirical studies on the effectiveness of disclosure limitation methods such as microaggregation, additive noise, and data swapping for official microdata were conducted by Ito and Murata (2011), Ito and Hoshino (2012, 2013, 2014). Empirical research on using top coding and recoding to create anonymized microdata that contain more detailed geographical information was conducted by Ito et al. (2015, 2016). Empirical research on adaptability of perturbative methods such as data swapping and PRAM (=Post RAndomization Methods) used to reduce disclosure risk for official microdata was conducted by Ito et al. (2017, 2018).

As part of preparations for the 2020 U.S. Population Census, the U.S. Census Bureau has developed a new methodology for the release of official statistical tables. This method creates privacy-preserved official microdata based on the concept of differential privacy. The adaptability of differential privacy to official statistical data in Japan is expected to become a topic in the future.

In this paper, we survey the current discussion about differential privacy used for creating and releasing official statistical data, including the discussion in the United States. We also investigate the possibility of adapting differential privacy for detailed geographical data from the Japanese Population Census, and examine the potential of differential privacy as an anonymization method for Japanese statistical data. In an era of big data and artificial intelligence, the risk of sophisticated disclosure attacks is increasing. If differential privacy can be applied to detailed geographical data for official statistics in Japan, it can potentially strengthen Japanese statistics against such attacks.

1 Secondary Use of Official Statistics in Japan

In most countries, official statistics are available in the form of statistical tables and microdata, the latter of which is provided in a variety of forms. Specific examples include the creation and provision of anonymized microdata (data created by applying anonymization methods to original microdata), managing access to original (non-anonymized) microdata, provision of tailor-made statistical tables, and remote execution services. While the specifics of how microdata are provided vary by country, they are usually provided in various forms according to users' needs and in consideration of confidentiality requirements under each country's legal system.

In Japan, two types of statistical data are provided: statistical tables (open data) and microdata created based on questionnaire information (original microdata). Specifically, provision of confidential data, creation and provision of anonymized

microdata, and tailor-made tabulation services are carried out under the Japanese Statistics Act (Article 33, Provision of Questionnaire Information; Article 34, Production of Statistics, etc. by Entrustment; Article 35, Production of Anonymized Data; Article 36, Provision of Anonymized Data).

Statistical tables in Japan are made available on paper and/or via the Internet through "e-Stat," which serves as a comprehensive portal for government statistics. For data from statistical surveys conducted by the Statistics Bureau of the Ministry of Internal Affairs and Communications, API (Application Programming Interface) and statistical GIS (Geographic Information System) functions have been added to further improve the convenience of accessing official statistics as open data. The Basic Act on the Advancement of Public and Private Sector Data Utilization enacted in December 2016 is promotes the use of open data for official statistics, administrative data and private "big data".

Meanwhile, the use of official statistical data, administrative data and private "big data" to promote evidence-based policymaking (EBPM) has been receiving increased attention in Japan. In 2017, the Council for the Promotion of Statistical Reform was established to discuss ways to create and provide official microdata, including anonymized microdata and original data, in an EBPM-oriented manner. Against this background, a revised Statistics Act was enacted on 1 June 2018 and came into force on 1 May 2019. One feature of the revised act is that it expands the scope for provision of questionnaire information. Specifically, clause 2 of Article 33 was newly added, and Article 36 was revised. The revisions state that the "significant public benefit" in Articles 33 and 36 should be of a level to which "the preparation of statistics, etc., does not compromise citizen confidence regarding official statistical surveys" and is "objectively reasonable and appropriate" (Legal Study Group on the Utilization and Provision of Questionnaire Information, Etc., 2018, p. 2). As a result, it is necessary to place usage purpose-based limitations on questionnaire information (original microdata) and anonymized microdata at the operational level.

Such advancement in the secondary use of official statistics may lead to the adoption of remote execution systems, and may also impact how official statistics results are published in Japan.

2 Differential Privacy Initiatives for Publication of Official Statistics

At present, the Statistics Bureau of Japan releases 6 types of anonymized microdata from Japanese official statistics. In order to create anonymized official microdata, perturbative and non-perturbative techniques are applied to official microdata. Anonymized microdata from the Population Census is made available, with anonymized microdata from the 2000 and 2005 Census currently available. Various disclosure limitation methods such as sampling (at a sampling rate of 1%), recoding, top (bottom)

coding, and data deletion are applied to the data before it is released. Data swapping is applied as an additional perturbative method to create anonymized Census microdata.

In order to increase the use of anonymized official microdata, several empirical studies on the effectiveness of disclosure limitation methods such as microaggregation, additive noise, and data swapping for official microdata were conducted by the National Statistics Center (Ito and Murata (2011), Ito and Hoshino (2012, 2013, 2014)). In preparation for the release of anonymized microdata from the 2010 Population Census, empirical research was conducted by the Statistics Bureau of Japan and the National Statistics Center (Ito et al. (2015, 2016, 2017, 2018)), while empirical research on using top coding and recoding with aim of creating anonymized microdata that contain more detailed geographical information was conducted by Ito et al. (2015, 2016). Empirical research on the potential of perturbative methods such as data swapping and PRAM (=Post RAndomization Methods) to reduce disclosure risk for official microdata was conducted by Ito et al. (2017, 2018).

In various countries, differential privacy methods have been proposed as a method for controlling noise according to the standards of secrecy. In the area of computer science, differential privacy has been developed from the concept of formal privacy.

Differential privacy (Dwork (2006)) is an indistinguishably-based definition of privacy where the output distribution of an differentially-private random algorithm (generally referred to as a "mechanism") from database D1 is almost identical to that from any neighbouring database (i.e., any database where only one record is different from D1), and thus can limit the disclosure of private information from an individual record contained in the database.

The closeness between the output distributions, usually denoted by ε, is the security parameter of differential privacy; smaller ε guarantees stronger privacy but tends to reduce the utility of the output, e.g., $\varepsilon=0$ guarantees perfect privacy but the output becomes completely useless, while $\varepsilon=\infty$ doesn't guarantee any privacy. Further details of the definition and characteristics of differential privacy can be found in Dwork (2006) and Dwork (2008).

Abowd (2018) described the possibility of database reconstruction attacks on statistical tables. A reconstruction attack exposes personal information contained in confidential data that is the source of a query by combining a small number of random queries, even without having to look carefully at the queries themselves (Dinir and Nissim, 2003). From this, the idea of differential privacy arose by determining an appropriate privacy-loss budget ε and returning the query with noise (Dwork, 2006).

Formal privacy methodologies have been discussed in the field of computer science. Rather than ad hoc application of confidentiality to individual information contained in confidential data, formal privacy starts with a mathematical definition of privacy, and a mechanism is considered for returning queries using confidential data in a form that is consistent with this formal privacy definition. By applying this concept, statistical result

tables are modelled as a series of queries applied to confidential data. Queries return data that is formally private data instead of confidential dataactual survey items. Based on these ideas, the potential of differential privacy for official statistics is being investigated.

Abowd (2018) showed that when many geographic categories are used to create detailed statistical tables, combining these tables can increase the risk of identifying individuals even if the tables do not contain identifying information. Avoiding such risks requires publishing secure statistical tables with added noise, but at the same time the numerical accuracy of the results needs to be maintained. To resolve this issue, on-demand systems such as the TableBuilder developed in Australia create statistical tables based on variables selected by the user. This is done by applying random noise according to the cell frequency.

In the United States, the U.S. Census Bureau is engaged in a major project that uses differential privacy. In preparation for the 2020 U.S. census, the U.S. Census Bureau is performing verifications using 2010 census data, and investigating the potential of differential privacy as a way of maintaining data accuracy, while ensuring data security for statistical tables containing nationwide data such as gender, race, age, and relation to head-of-household. Specifically, because statistical tables can be created at the state, county, census tract, or block level, a privacy-loss budget ε is set, and an optimal value for ε is determined based on trade-offs between privacy loss and accuracy. When creating differentially private statistical tables by adding noise based on a mathematically optimized privacy-loss budget ε, a state-level geographical category is assigned to microdata corresponding to the individual unit. Statistical tables at the county, census tract, and block levels are created in the same way, with small area microdata similarly assigned at the corresponding level. In this way, microdata corresponding to the confidential data in statistical tables are newly created with pseudo-regional classifications at state, county, census tract, census tract, or block level.

To examine the potential of differential privacy as an anonymization method, the U.S. Census Bureau is applying visualization of optimal values for privacy loss and data accuracy based on the concept of the "production probability frontier" from economics. By plotting the estimated marginal social benefit curve, it becomes possible to determine ε such that marginal social benefit and marginal social cost agree. Also, when applying differential privacy methodologies, it is necessary to eliminate the trade-off relationship between the various positions on the user and creator sides of the population census.

Differential privacy has so far been rarely discussed in the context of Japanese official statistics, whereas in other countries, discussions of differential privacy in the field of official statistics have progressed. The process undertaken by the U.S. Census Bureau, which applies differential privacy to high-dimensional cross-tables at sub-regional levels and tries to expand this into higher-level regional level aggregation result tables for publishing, may be an approach worth considering when it comes to exploring the potential adoption of differential privacy in Japan.

3 Experiments on the Application of Differential Privacy to Small-Area Statistics

In order to explore the potential of differential privacy for official statistics in Japan, we applied differential privacy to meshed population data from the Japanese Population Census. Since differential privacy is a definition, not an algorithm, there will be many differential privacy algorithms with ε for achieving this task; the utility of their outputs may differ considerably (Dwork and Roth (2014)) , and thus it is important to find a differential privacy algorithm that can achieve smaller ε while preserving better utility of its output.

The simplest method to apply differential privacy to the meshed population data would be to add random values following a double exponential distribution (a Laplace distribution) to all population values in the mesh (including "nonstructural zeros"). This method is known as the "Laplace mechanism" (Dwork et al. (2006)).

Here, the value of the privacy loss budget ε for differential privacy security is determined according to the scale of the Laplace noise to be added. Specifically, letting Lap(λ) be the Laplace noise at scale λ for a given mesh population value p_i, the population values with added Laplace noise $p_i' = p_i + \mathrm{Lap}(\lambda)$ satisfy differential privacy $\varepsilon = 1/\lambda$ also written $(1/\lambda)$ differential privacy.

However, it is difficult for output obtained from a simple Laplace mechanism such as this to be utilized as statistical data. Three issues in particular have emerged in past research on the subject.

1. Violation of nonnegativity constraints. When using the Laplace mechanism, output can contain many negative values that would be impossible in actual population data.

2. Inflation of data amounts. Nearly all cells that contained nonstructural zeros will come to hold nonzero values, which will significantly increase the amount of output data when applied to large-scale sparse data such as census data.

3. Degraded partial-sum accuracy. Because noise is uniformly added to each cell, error when taking partial sums of multiple cells will increase, which in turn decreases data accuracy.

Deviation from nonnegativity constraints can be simply addressed by replacing all cells containing negative values with 0, in other words $p_i' = \max(p_i + \mathrm{Lap}(\lambda), 0)$. Repeated application of this procedure will increase the number of cells containing 0, thereby alleviating increases in data quantity. However, this is equivalent to applying a positive bias to the overall dataset, further exacerbating the problem of partial-sum accuracy.

A method based on nonnegative wavelet transformation with top-down refinement has been proposed as an attempt to solve the problems described above Terada et al. (2015). In reference to the "privelet" by Xiao et al. (2011), this method applies a Haar wavelet transform to the original population data, and Laplace noise is added to the resulting wavelet coefficients. These coefficients are subjected to top-down inverse wavelet

transform, correcting their values so that the output does not violate nonnegativity constraints, thereby providing population data that is in accordance with differential privacy.

In order to handle two-dimensional data such as meshed population data, the method introduces a Morton order map (Morton (1966)) (a type of locality-preserving map) for transforming the data into one-dimensional data before applying Haar wavelet transform, rather than use multi-dimensional wavelets as in the case of privelets. The noise scale needed for multi-dimensional wavelet coefficients is much larger than that of one-dimensional wavelet coefficients for the same ε, and therefore introducing the Morton order map to avoid involving multi-dimensional wavelets preventing increases in noise strength resulting from use of multi-dimensional wavelets.

By obtaining population data in this way, the characteristics of wavelet transformation control the accuracy of partial sums without violating nonnegativity constraints. Further, the process for correcting wavelet coefficients to avoid violating nonnegativity constraints restores data sparseness. We can thus expect improvements for each of the three issues described above.

We applied this method to data from the 2010 Population Census for a 1/2 standard regional mesh (n=512×512=262,144 meshes, 500m per side) over a 256-km^2 area in the Tokyo region. The following describes the results as compared with those from applying the Laplace mechanism and the Laplace mechanism with nonnegativity correction.

Figure 1 shows the original data, with darker colors indicating higher populations. The data contained 95,317 nonzero values, for a density ratio of approximately 36%. This is a remarkably high density for a Census dataset as the region includes areas with the highest population densities in Japan. However, it also includes uninhabitable areas (lakes and marine areas).

Figure 2 contains the results from the Laplace mechanism (ε=0.1). Regions shown in red are those violating nonnegativity constraints (taking negative values), accounting for 88,399 (approximately 1/3) of all cells. All cell values are nonzero, so density is 100%. This reflects the nature of the Laplace mechanism, in which the sparsity of output data will be almost 100% lost, regardless of the sparsity of the original data.

Figure 3 shows the results from the Laplace mechanism with nonnegativity correction (ε=0.1). This is equivalent to the output in Figure 2 when replacing negative values with 0, thereby removing all red regions. Because all negative values in the output (values for 88,399 cells) are replaced with zeros, density decreased to approximately 64%. However, this still represents an approximate doubling of the density compared to the original data.

Figure 4 shows the results of applying the method in Terada et al. (2015), with ε=0.1. There are no red regions indicating violation of nonnegativity constraints, so this issue is resolved. Further, sparsity is approximately 28%, so sparsity of the original data is largely maintained.

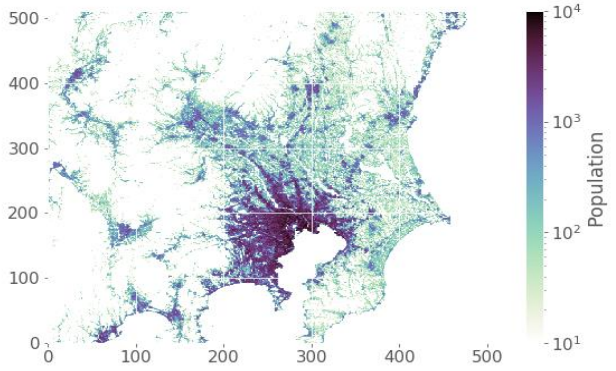


Fig 1 Results based on the original data.

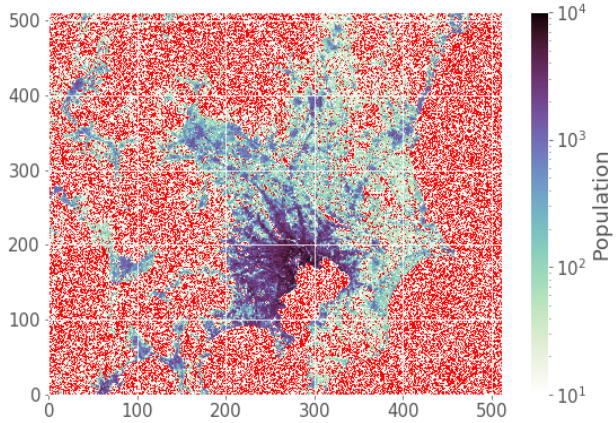


Fig 2 Results based on the Laplace mechanism, ε=0.1.

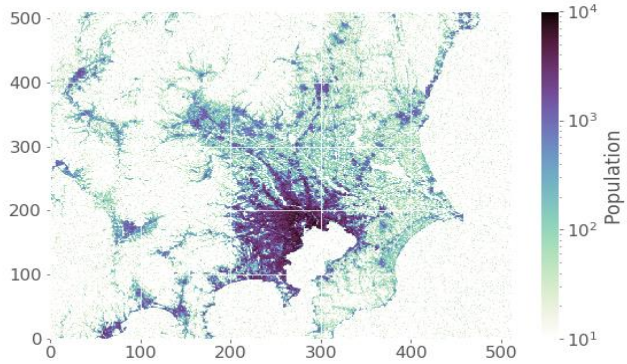


Fig 3 Results based on the Laplace mechanism with nonnegativity correction, ε=0.1.

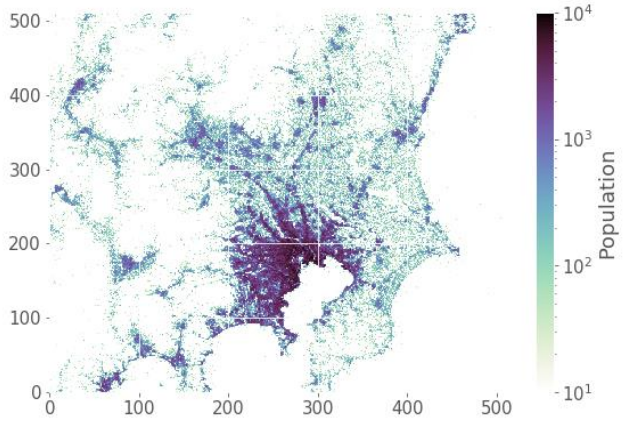


Fig 4 Results of applying the method in Terada et al. (2015), with ε=0.1.

Regarding partial sums error in the output for each method, we changed the privacy loss budget ε for differential privacy security to values in {0.1, 0.2, ln 2, ln 3, 3, 5} and compared the results (See Appendix Figures 1, 2 and 3).

Tables 1, 2 and 3 contain the size of error (mean absolute error over 100 trials) for partial sums and changing ε values under the Laplace mechanism, the Laplace mechanism with nonnegativity correction, and the method of Terada et al. (2015), respectively. The top rows in these tables contain the area of the partial sum. For example, values in the "16^2" column indicate mean error for the population in a square region with approximately 16-km sides (note that area sizes are approximate because the mesh sizes slightly vary with latitude).

Under the Laplace mechanism, the error is small for partial sums over small regions, but greatly increases for larger regions. Further, results retain the above-described issues of violated constraints and increased data quantity. The Laplace mechanism with nonnegativity correction does not violate constraints, and the increase in data is somewhat mitigated, but as the area of partial sums becomes larger, the error becomes larger than under the normal Laplace mechanism. For example, for the largest partial sum calculated in this experiment, that for a 256-km^2 region (the full area considered in this experiment), the error in output from the Laplace mechanism with nonnegativity correction is over 100-fold that of the error in output from the normal Laplace mechanism for all values of ε. In other words, not only has the issue of degraded partial-sum accuracy not improved, it has become far worse.

In contrast, under the method of Terada et al. (2015), the error increase accompanying expansion of the partial sum region is much better controlled. In cases of small areas (approximately 1 km^2, etc.), there is relatively large error in comparison with the other

Area size(km^2)	1^2	2^2	4^2	8^2	16^2	32^2	64^2	128^2	256^2
ε=0.1	21.9	44.8	90.2	180.4	362.1	738.9	1483.3	3068.9	6547.7
ε=0.2	10.9	22.4	45.0	90.4	181.3	363.7	723.9	1389.5	2807.1
ε=ln2	3.2	6.5	13.0	26.0	51.7	104.1	208.8	415.3	865.7
ε=ln3	2.0	4.1	8.2	16.4	32.9	65.4	131.0	282.0	625.0
ε=3	0.7	1.5	3.0	6.0	12.0	23.8	48.8	102.7	209.0
ε=5	0.4	0.9	1.8	3.6	7.2	14.4	28.9	59.4	131.6

Table 1 Output error under the Laplace mechanism.

Area size(km^2)	1^2	2^2	4^2	8^2	16^2	32^2	64^2	128^2	256^2
ε=0.1	20.1	64.5	229.4	876.9	3464.1	13824.3	55291.9	221167.5	884669.9
ε=0.2	10.1	32.0	112.7	428.2	1686.6	6728.5	26910.5	107641.9	430567.6
ε=ln2	2.9	9.2	32.1	121.1	475.3	1895.0	7578.1	30312.5	121249.8
ε=ln3	1.9	5.8	20.2	76.1	298.8	1190.4	4760.4	19041.7	76166.6
ε=3	0.7	2.1	7.4	27.8	109.1	434.6	1738.0	6952.0	27807.8
ε=5	0.4	1.3	4.4	16.7	65.5	260.9	1043.2	4172.7	16690.8

Table 2 Output error under the Laplace mechanism with nonnegativity correction.

Area size(km^2)	1^2	2^2	4^2	8^2	16^2	32^2	64^2	128^2	256^2
ε=0.1	44.5	60.1	75.0	88.3	100.1	104.3	106.3	119.7	147.8
ε=0.2	24.6	31.9	38.8	44.8	49.4	53.8	60.2	64.3	109.4
ε=ln2	7.9	9.8	11.6	13.1	14.5	15.7	16.6	17.2	23.4
ε=ln3	5.1	6.3	7.4	8.3	9.1	9.9	10.5	11.3	16.6
ε=3	1.9	2.4	2.7	3.1	3.4	3.6	3.9	4.4	6.3
ε=5	1.2	1.4	1.6	1.9	2.0	2.2	2.4	2.5	3.7

Table 3 Output error under the method of Terada et al. (2015).

two methods, but the error increases gradually as the size of the area increases. This tendency holds for all values of ε, so the mean error for a 256 km^2 region is only approximately 150 people when ε=0.1, and approximately 4 people when ε=5. This is only 1/40 to 1/30 the error under the normal Laplace mechanism, and 1/6,000 to 1/4,000 the error under the Laplace mechanism with nonnegativity correction. This demonstrates

that the method proposed in Terada et al. (2015) avoids constraint violations and increases in data, and furthermore addresses the issue of degraded partial-sum accuracy.

4 Conclusion and Future Developments in Official Statistics

This paper examined the current state of secondary use of official statistics and the applicability of differential privacy for data anonymization in Japan. Our experiment using Japanese Census data demonstrated that applying noise to statistical tables based on the concept of differential privacy results in changes of cell values according to the value of ε. Results show that identical values of ε can result in different degrees of noise addition depending on the underlying mechanism used for applying differential privacy. While the adoption of differential privacy for official statistics in Japan could be viewed as challenging, it is worthwhile to consider the concept of differential privacy as developed in the field of computer science in order to further advance developments in official statistics in Japan.

References

Abowd, J. M. (2018) "Staring-Down the Database Reconstruction Theorem", presented at Joint Statistical Meetings, Vancouver, BC, Canada.

Dinur, I., Nissim, K., (2003) "Revealing Information while Preserving Privacy" in Proceedings of Twenty-Second *ACM SIGMOD-SIGACT-SIGART symposium on Principles of Database Systems* (PODS '03), ACM, NEW York, NY, USA.

Dwork, C. (2006) "Differential Privacy," in Proc. 33rd intl. conf. Automata, Languages and Programming - Volume Part II, pp. 1–12.

Dwork, C., McSherry, F., Nissim, K., Smith, A. (2006) "Calibrating Noise to Sensitivity in Private Data Analysis," in *Proc. 3rd conf. Theory of Cryptography*, pp. 265–284.

Dwork, C. (2008) "Differential Privacy: a Survey of Results," in Proc. 5th intl. conf. Theory and applications of models of computation, pp. 1–19.

Dwork, C. and Roth, A. (2014) "The algorithmic foundations of differential privacy," Foundations and Trends in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211–407,.

Ito, S. and Murata, M. (2011) Quantitative Methods to Assess Data Confidentiality and Data Utility for Microdata in Japan, Paper presented at Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Tarragona, Spain, pp.1-10.

Ito, S. and Hoshino, N. (2012) The Potential of Data Swapping as a Disclosure Limitation Method for Official Microdata in Japan: an Empirical Study to Assess Data Utility and Disclosure Risk for Census Microdata, Paper presented at Privacy in Statistical Databases 2012, Palermo, Sicily, Italy, pp.1-13.

Ito, S. and Hoshino, N. (2013) Assessing the Effectiveness of Disclosure Limitation Methods for Census Microdata in Japan, Paper presented at Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Ottawa, Canada, pp.1-10.

Ito, S. and Hoshino, N. (2014) Data Swapping as a More Efficient Tool to Create Anonymized Census Microdata in Japan, Paper presented at Privacy in Statistical Databases 2014, Ibiza, Spain, pp.1-14

Ito, S., Hoshino, N., Akutsu, F. (2015) A Quantitative Assessment of Data Confidentiality and Data Utility to Create Anonymized Census microdata in Japan, Paper presented at Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Helsinki, Finland, pp. 1-14.

Ito, S., Hoshino, N., Akutsu, F. (2016) Potential of Disclosure Limitation Methods for Census Microdata in Japan, Paper presented at Privacy in Statistical Databases 2016, Dubrovnik, Croatia, pp.1-14.

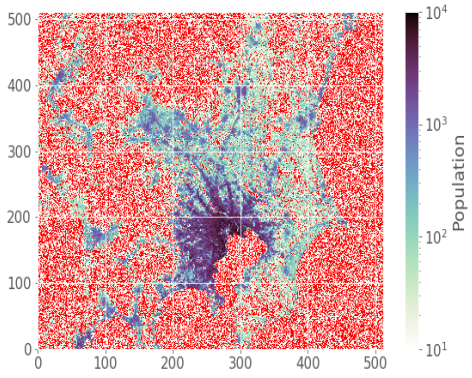
Ito, S., Hoshino, N., Akutsu, F., Kikuchi, R. (2017) Investigating New Methods for Creating Anonymized Microdata Based on Japanese Census Data, Paper presented at Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Ministry of Foreign Affairs, Skopje, Macedonia, 2017, pp.1-16.

Ito, S., Yoshitake, T., Kikuchi, R., Akutsu, F. (2018) "Comparative Study of the Effectiveness of Perturbative Methods for Creating Official Microdata in Japan" Josep Domingo-Ferrer and Francisco Montes (eds.) *Privacy in Statistical Databases: UNESCO Chair in Data Privacy, International Conference, PSD 2018, Valencia, Spain, September 26–28, 2018, Proceedings (Lecture Notes in Computer Science)*, Springer, pp.200-214.

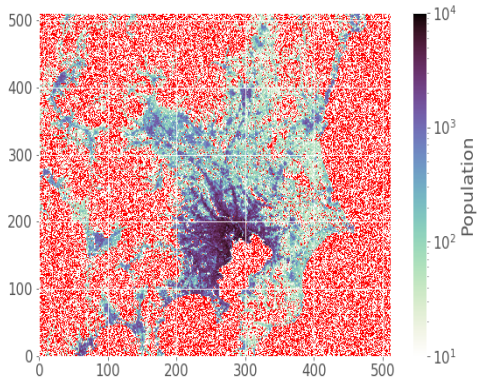
Morton, G. M. (1966) "A Computer Oriented Geodetic Data Base; and a New Technique in File Sequencing," International Business Machines Co. Ltd, Canada, Technical Report.

Terada, M., Suzuki, R., Yamaguchi, T. and Hongo, S. (2015) On Publishing Large Tabular Data with Differential Privacy (in Japanese), *IPSJ J.*, 56(9), pp.1801-1816, Information Processing Society of Japan.

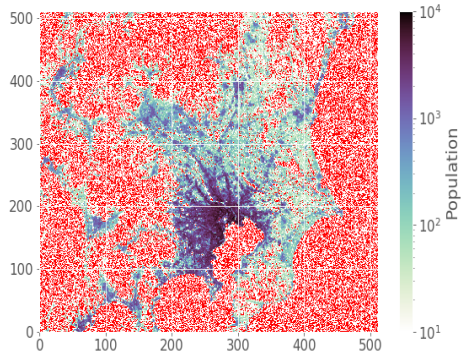
Xiao, X., Wang, G., Gehrke, J. and Jefferson, T. (2011) Differential Privacy via Wavelet Transforms, *IEEE Trans. Knowledge and Data Engineering*, 23(8), pp.1200–1214, IEEE.



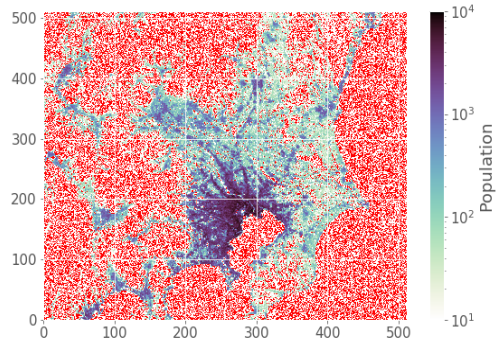
ε=0.1



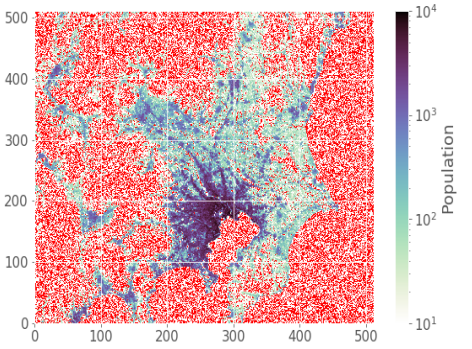
ε=0.2



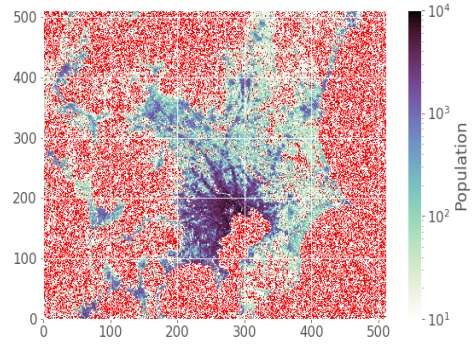
ε=ln 2



ε=ln 3

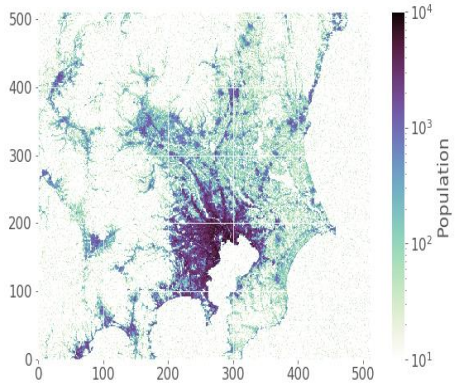


ε=3

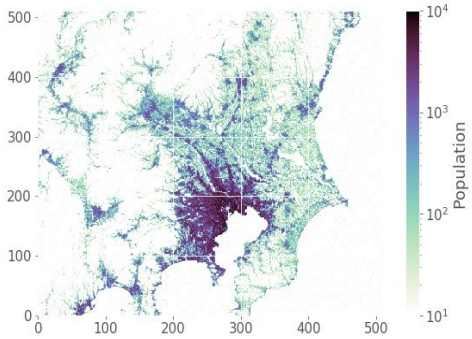


ε=5

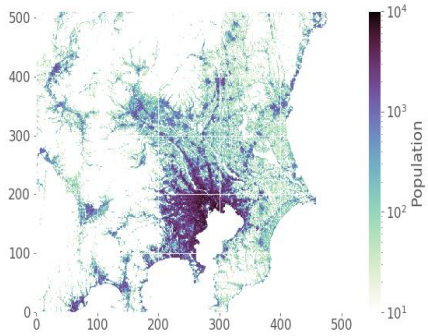
Appendix Figure 1 Comparison of Results from the Laplace mechanism, ε=0.1, 0.2, ln 2, ln 3, 3, 5.



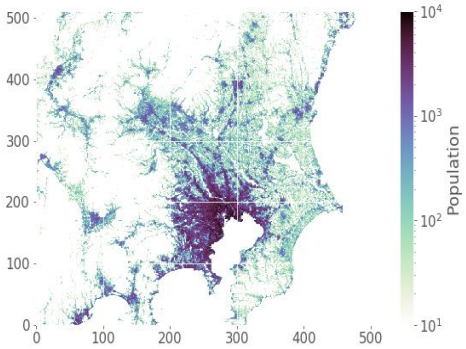
ε=0.1



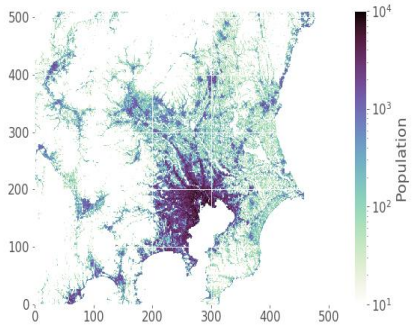
ε=0.2



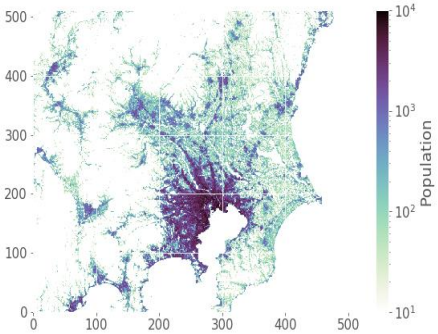
ε=ln 2



ε=ln 3

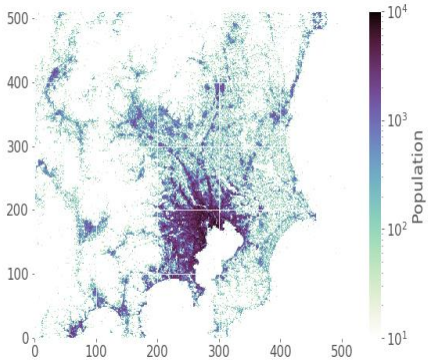


ε=3

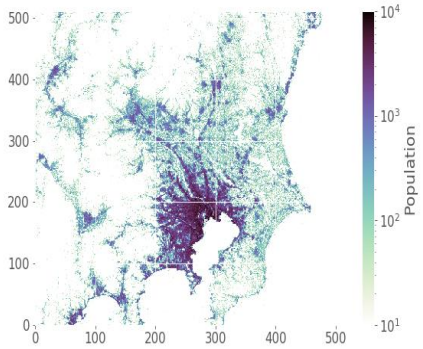


ε=5

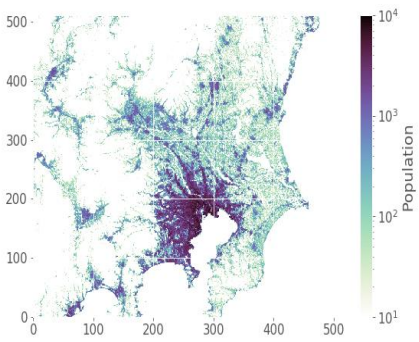
Appendix Figure 2 Comparison of Results from the Laplace mechanism with nonnegativity correction, ε=0.1, 0.2, ln 2, ln 3, 3, 5.



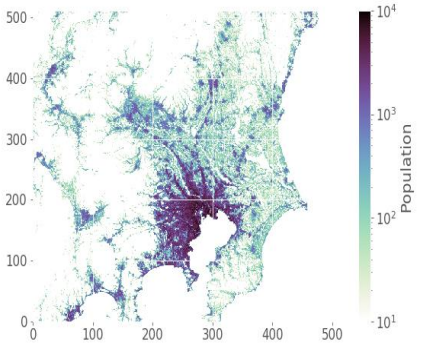
ε=0.1



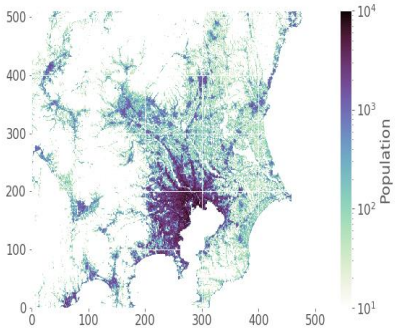
ε=0.2



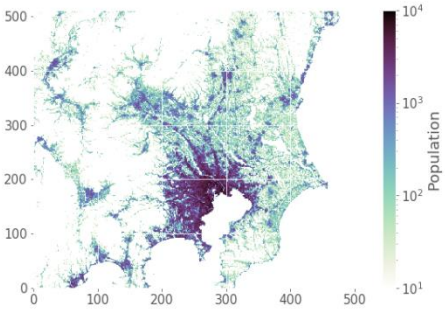
ε=ln 2



ε=ln 3



ε=3



ε=5

Appendix Figure 3 Comparison of Results of applying the method in Terada et al. (2015), ε=0.1, 0.2, ln 2, ln 3, 3, 5.