Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 219 (2023) 1120–1127

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2022

Web-tool for optimizing locations of health centers

Pasi Fränti[a]*, Radu Mariescu-Istodor[b], Awais Akram[a]

[a]*University of Eastern Finland, Joensuu, Finland*
[b]*Karelia University of Applied Sciences, Joensuu, Finland*

Abstract

Coronary heart disease patients must reach a hospital within 90 minutes from the start of myocardial infarction to avoid the risk of death. Current hospital locations are optimized for the overall health care system and may not be the best possible for these patients. For this reason, we have constructed a web tool for optimizing the facility locations to study how well they serve for the coronary heart disease patients. In this paper, we explain how the system has been constructed, how it can be used, and show the results using patient data in Finland. Results show that the number of patients at risk could theoretically be reduced to 1% with optimized locations but at the cost of increasing the average travel time.

© 2023 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2022

Keywords: Health care information systems; facility optimization; clustering; coronary heart disease; web tools

* Corresponding author.
E-mail address: pasi.franti@uef.fi

1877-0509 © 2023 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2022
10.1016/j.procs.2023.01.392

1. Introduction

Myocardial infarction patients need to be treated within a given time limit or they face potentially fatal consequences [1–2]. Patients living closer to a health center have higher change of survival than those who live far away [3]. Current health care system is under pressure to cut costs and optimize the health care services better. This may lead to reducing the number of health centers providing treatment for myocardial infarction patients.

In this paper, we present web-tool (Fig. 1) to analyze how well the current health station locations serve for the people using computer simulation. The method involves facility location optimization with real patient data and their location with the accuracy of postal code. We explain the components of the system, what data is needed, how it is used, and how the optimization is done. We demonstrate its performance by studying the results on Nationwide health care data collected during 2015-2018.

The optimization tool repositioning the health centers with one of the four alternative optimization criteria: (1) Euclidean distance, (2) travel distance, (3) travel time, (4) the number of patients at risk by. The optimization is based on clustering with travel time constraint. We demonstrate how the different parameters affect the proportion of patients within safe distance to the hospital versus the patients at risk. Experiments show that 5% of the patients are at risk with the current locations while it could theoretically be reduced to 1% with optimized locations.

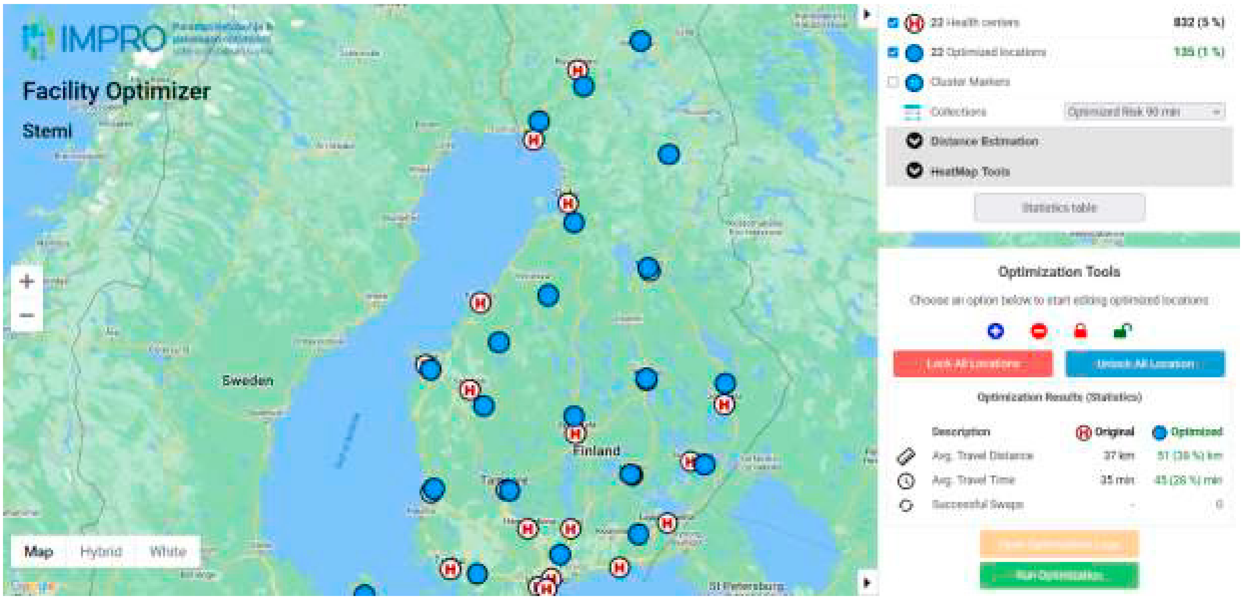

Fig. 1. Web-tool facility location optimizer for myocardial infarction patients in Finland.

2. Web tool

The system consists of the following main components:

- Patient database
- Patient locations
- Travel time estimation
- Clustering component
- Web-interface

2.1. Patient database

The dataset represents the location of 17,563 adult patients who suffered acute myocardial infarction during the years 2015-2018. The data was extracted from the national electronic patient database which contains information on 4,280,985 patients (78% of Finnish population) collected during years 2015-2018. The patients were selected to include those that had some ICD10 [4] diagnosis in the range of I21.0-I21.3. This includes all types of myocardial infarction except those of unspecified type and those with Non-ST elevation (NSTEMI) which are not as critical.

2.2. Patient locations

The location of patients is known as postal codes which we converted to GPS locations. We used a 1km × 1km population density grid information from the National Land Survey of Finland. The GPS coordinate was calculated as the weighted average of the centre points of all blocks that are located within the postal code area. There are 3038 different postal codes in Finland which are larger in sparsely populated areas.

For privacy reasons the exact location of patients would be difficult to obtain, but the postal code precision is good enough in this scenario as the errors from the estimation are minor compared to the 90-minute hospital reachability threshold.

2.3. Travel time estimation

Sometimes it is enough to use just geographical (Euclidean) distance and then estimate the travel time using some default speed. This is reasonable especially when the landscape is flat and road network relatively straightforward [5]. However, the quality of such estimation can degrade a lot in areas with natural borders like rivers, lakes, forest and canyons. In large part of Finland, there are lots of lakes that affect the road network. Other affecting factors include speed limitations, traffic congestions and lack of roads in rural areas.

Ideally, one would calculate the shortest path between two locations using a road-network and path-finding algorithms such as Floyd, Dijkstra or A* [6]. However, the optimization involves huge number of calculations, and these becomes simply too slow. In our case, the optimization would take hours or even days while the system is designed to be, if not real-time, at least interactive with small waiting times.

Another solution would be to use some open-source routing engines like *Open-Source Routing Machine* (OSRM) which provides API for computing the fastest path, and also a batch functionality with their so-called table service or distance matrix APIs. However, the facility locations change dynamically during the optimization process so single call of such engine would not be enough.

In this paper, we use the *overhead graph* [7], see Fig. 2. It creates a complete graph where a subset of locations are the nodes. We then calculate the distance between all nodes using both Euclidean and road network distance. The ratio between these two distances is called *overhead*, and it estimates the error if Euclidean distance is used between two locations close to the nodes. During the optimization, the distance between a patient and health centre are calculated based on their Euclidean distance multiplied the weight between the nodes in the graph closest to the patient and health centre locations. This gives significantly more accurate estimation (on average) than simple Euclidean distance and takes only fraction of time what shortest path calculation would require.

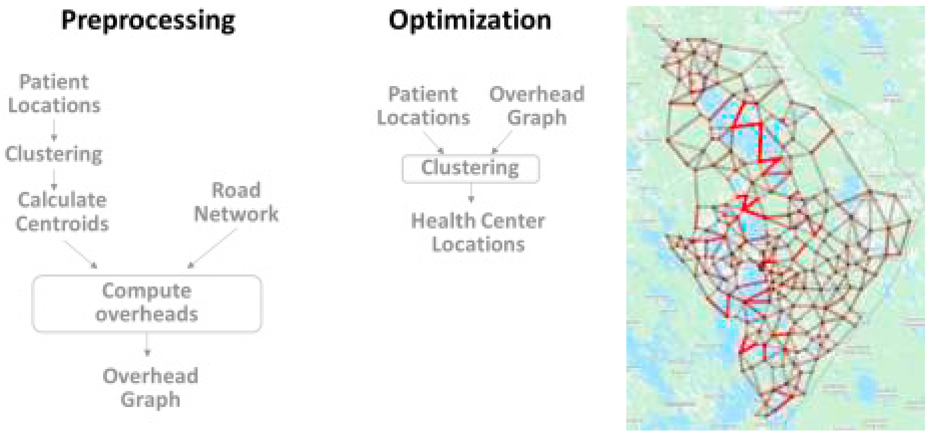

Fig. 2. Use of the overhead graph for the trave time calculations (left). A sample graph in North Karelian region (right). The intensity of red indicates the weight of the edges. Larger weights typically appear near lake where distance via road network is higher than Euclidean distance.

2.4. Clustering component

Optimization is based on k-means clustering with few differences. First, k-means works reasonably well when the clusters overlap, and the algorithm is repeated several times [8]. However, to make sure the optimization is not degraded by inferior clustering result, we used *random swap* algorithm which is slight extension of k-means by two additional components: centroid swap and selection [9].

The second difference is that we do not want to minimize Euclidean distance but the real travel distance between patients and the nearest hospital. Four optimization goals are implemented in the clustering algorithm: (1) Euclidean distance, (2) travel distance, (3) travel time, and (4) the number of patients at risk. The algorithm is the same in all these three cases, but the selection of the nearest hospital depends on this choice, see Fig. 3. In case of travel distance and travel time, we use the result provided by the overhead graph either as such, or by converting it to the estimated travel time.

In case of minimizing patients at risk, we need different function in the partition step. Obvious solution would be to implement simple binary function giving penalty 0 for patients within the critical 90 minutes travel time, and penalty 1 for others. However, such discrete (binary) objective function can lead to unwanted effects in the optimization process, and for this reason, we use Sigmoid function (Fig. 4) which optimizes for the same goal but is smoothed version of its binary counterpart.

```
K-Means (P, K) → (C, L)
  Input : points P = {p1,...,pN}
          the number of clusters K
  Output: cluster centers C = {c1,...,cK}
          cluster labels  L = {label(i), i=1,...,N}

  Randomly choose K initial centers C = {c1,...,cK}
  REPEAT
  Cprevious ← C
  FOR all i ∈ [1, N] DO   // Partitioning
     label(i) ← arg min d(pi, cj)
  FOR all j ∈ [1, k] DO   // Centroid update
     cj ← Average of pi, whose label(i) = j
  UNTIL C = Cprevious
```

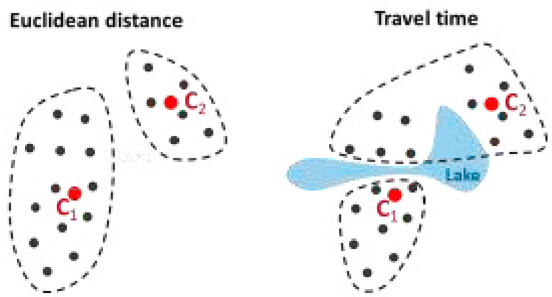

Fig. 3. The choice of optimization goal affects only the distance function *d* in the partition step of k-means (left). However, its effect on the partition can be huge significant in the areas where road network is not straightforward (right).

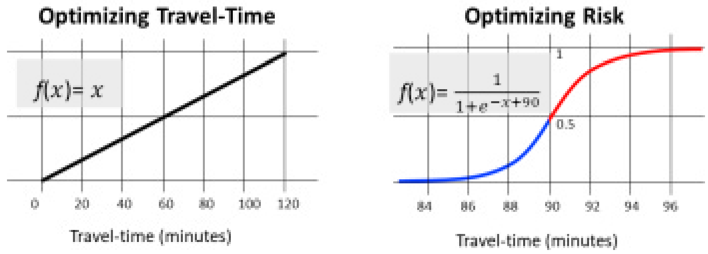

Fig. 4. The difference of optimizing travel time (left) and optimizing patients at risk (right).

2.5. Web interface

The system is wrapped up by sophisticated web interface as was shown in Fig. 1. It allows to change various parameters such as objective function, risk threshold, method for optimization goal, and the number of health stations. It is also possible to lock certain health stations and optimize the location of the others, see Fig. 5. The number of stations can also be manually tuned. Optimization of most common combinations are pre-calculated and stored in the system.

For the optimization goal, we consider the travel distance, travel time and the patients at risk. For the risk criterion, five different thresholds are considered: 30, 45, 60, 75 and 90 mins. Optimization is always started from the original locations of the health centres, but it is also possible to use a subset of the five university hospitals (*Helsinki*, *Turku*, *Tampere*, *Kuopio*, *Oulu*), or remove two stations (*Suupohja* and *Savonlinna*) with the least number of patients and therefore speculated to be at risk for removal.

The optimization results for all these combinations have been pre-calculated and can be examined in the system real-time. Other combinations can also be tested but the optimization may take few minutes, so it is not real-time and not necessarily interactive (<10s). The result of the optimization is shown at map, and by giving a brief statistical summary of the average travel time and distance, and the number of patients at risk. The choices of user-selected parameters are summarized in Fig. 6.

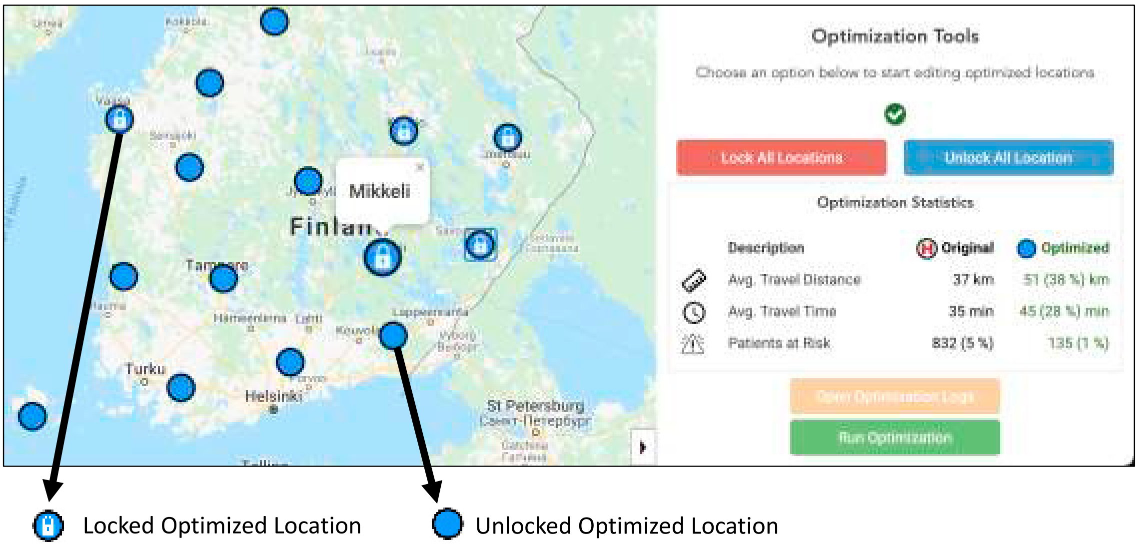

Fig. 5. Possibility to lock certain locations while optimizing the others.

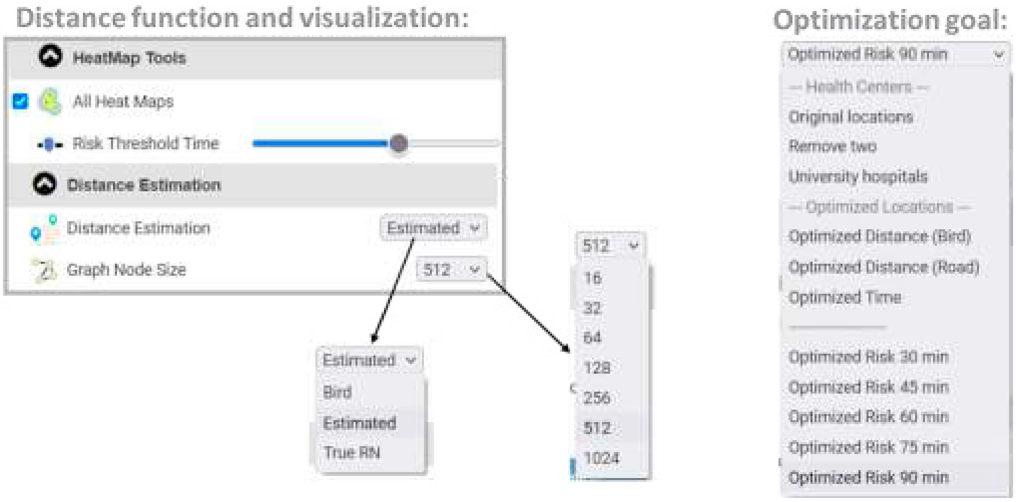

Fig. 6. User-selected parameters for the distance function, visualization and for the optimization goal.

3. Experiments

Optimization results are summarized in Table 1 using the four different optimization goals. Results are reported using the same attributes as used as the optimization goals. Results for the current locations are also provided for reference purpose. From the results we make the following observations. First, the best result for each attribute is always achieved by using the same attribute as the optimization goal. The choice of this goal is therefore the most important choice when using the tool in real life.

The second observation is that the number of patients at risk can be improved most. With 90 minutes risk threshold, the number patients at risk is 5% with the original locations. The optimized locations can reduce it from 832 (5%) to 135 (1%). The difference is big. Results when optimization for the patients at risk are further demonstrated in Fig. 7 with three risk thresholds (30 min, 60 min and 90 min). Reducing the risk threshold from 90 to 60 minutes would reach the same level (5%) as the current location with 90 minutes threshold. This means that the same number of people can reach the hospital in 1 hour time with optimized locations compared to 1 h 30 min.

The results also show that optimizing for the patients at risk would move many of the optimized centers to the sparsely population especially in North Finland and especially when the higher 90 min threshold is used. This significantly reduces the number of patients at risk but has the side-effect of increasing the average travel time for all patients. In other words, the optimization does not care if people need to travel 49 minutes (on average) instead of 35 minutes, as long as the time remains below 90 minutes, see Table 1.

Using travel distance or travel time as the optimization goal can also provide better locations and reduce those attributes, but the difference is not as significant. Optimizing travel distance would reduce by 5% from 36.6 km to 34.7 km, whereas optimizing travel time would reduce it by 3.7% from 35.3 min to 34.0 min. This would also reduce the number of patients at risk to 519 (distance) and 488 (time), respectively.

Table 1. Average travel time, distances, and the number of people at risk with the current and optimized locations. Boldfaces shows the cases when the particular attribute was directly optimized for.

Optimization goal:	Euclidean distance	Travel distance	Travel time	Patients at risk
Original locations	29.0 km	36.6 km	35.3 min	832 (5%)
Euclidean distance	**27.9 km**	36.1 km	36.2 min	792 (5%)
Travel distance	28.4 km	**34.7 km**	34.1 min	519 (3%)
Travel time (linear)	29.8 km	36.8 km	**34.0 min**	488 (3%)
Travel time (Sigmoid)	44.3 km	54.9 km	48.6 min	**135 (1%)**

4. Conclusions

We have presented a web-tool for optimizing health center locations using clustering algorithm and by minimizing the patients at risk. The tool can be very useful to simulate the effect of various decision like complete re-optimization of the locations or removing selected health stations. The results also showed that optimization based on a single factor like patients at risk can increase the average burden of others. A proper optimization criterion should therefore take consider more factors. It is also important to notice that the health centers do not serve only the myocardial infarction patients but have many other operations. Optimizing the locations merely for these is not feasible. However, considering the needs of the whole health care system would make the optimization goal very complex to define [10]. Nevertheless, that should be the future goal. Our study has demonstrated the power of a good web-tool for finding important insight that can help decision makers.

Acknowledgements

The project was funded by the Strategic Research Council (SRC) at the Academy of Finland (grant numbers 312703, 312706 and 336325).

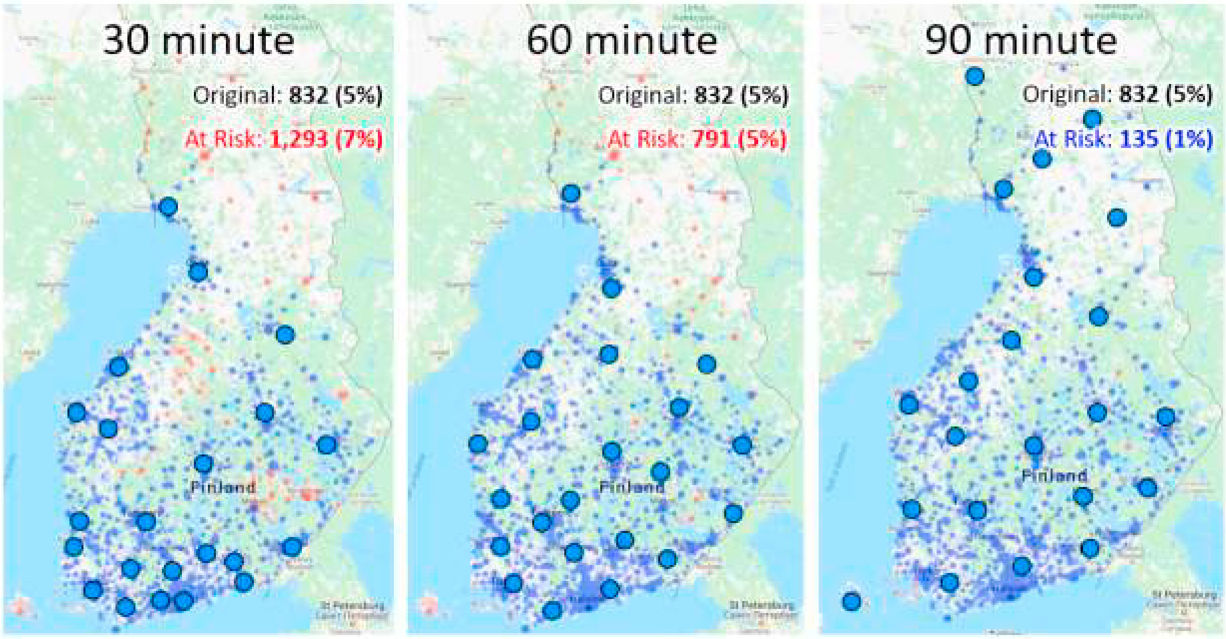

Fig. 7. Results of the optimization with three different thresholds. The colors of the locations are red for the patients at risk when using 30 min (left), 60 min (middle) and 90 min (right) thresholds. Otherwise, the color of the location is blue.

References

[1] Terkelsen JA et al. (2010). System delay and mortality among patients with STEMI treated with primary percutaneous coronary intervention. *JAMA*, 763-771.

[2] Berlin C, Panczak R, Hasler R, et al. (2016). Do acute myocardial infarction and stroke mortality vary by distance to hospitals in Switzerland? Results from the Swiss National Cohort Study. *BMJ Open*;

[3] Kelly C, Hulme C, Farragher T, & Clarke G (2016). Are differences in travel time or distance to healthcare for adults in global north countries associated with an impact on health outcomes? A systematic review. *BMJ open*, 6 (11).

[4] WHO (2016), *International statistical classification of diseases and related health problems, 10th revision*, Geneva: World Health Organization.

[5] Boscoe, FP, Henry, KA, & Zdeb, MS (2012). A nationwide comparison of driving distance versus straight-line distance to hospitals. *The Professional Geographer*, 64 (2), 188-196.

[6] Thomas H. Cormen (2009). Introduction to Algorithms. MIT Press.

[7] Mariescu-Istodor R and Fränti P (2021) Fast travel distance estimation using overhead graph, *Journal of Location-Based Services*.

[8] Fränti P and Sieranoja S, How much k-means can be improved by using better initialization and repeats?, *Pattern Recognition*, 93, 95-112, 2019.

[9] Fränti P (2018) Efficiency of random swap clustering, *J Big Data*, 5:13, 1-29.

[10] Wang, F. (2020). Why public health needs GIS: a methodological overview. *Annals of GIS*, *26*(1), 1-12.