Chapter 3
Citizen Science: A Case of Traffic Noise Measuring

Ivan Jezdović
Faculty of Organizational Sciences, University of Belgrade, Serbia
Zorica Bogdanović
Faculty of Organizational Sciences, University of Belgrade, Serbia
Bogdan Savković
Faculty of Organizational Sciences, University of Belgrade, Serbia
Dušan Barać
Faculty of Organizational Sciences, University of Belgrade, Serbia
Filip Filipović
Faculty of Organizational Sciences, University of Belgrade, Serbia

ABSTRACT

This paper discusses citizen science with particular focus on traffic noise measuring. The goal is to develop an approach that would enable citizens to collect traffic noise data through crowdsensing. In the experimental part, traffic noise data was collected during period of two weeks in Serbia's capital Belgrade, using a combination of participatory and opportunistic mobile crowdsensing. The obtained data was analyzed using statistical methods principal components analysis and cluster analysis. The obtained results identify traffic noise clusters of microlocations in the city and can be used for decision support for both citizens and city government.

Keywords: citizen science, traffic noise measuring, crowdsensing

INTRODUCTION

Noise pollution has a significant impact on the health and quality of life of people, especially in large cities. Accordingly, measuring noise and reducing its impact are important issues that are regulated in many countries. In the European Union, Directive 2002/49/EC requires cities with more than 100,000 inhabitants to construct noise maps and define plans for limiting and eliminating noise. Also, city authorities are obliged to communicate with citizens on topic such as environmental quality and noise pollution (*Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise - Declaration by the Commission in the Conciliation Committee on the Directive relating to the assessment and management of environmental noise - Public*, n.d.).

In addition to the classic way of generating maps (EEA(Europeon Environment Agency) et al., 2014), there is another approach allowed by the European Union directive and based on noise measurement. This approach requires a large number of measurements that is achievable thanks to the large number of smartphone users and the continuous improvement of the metrological aspects of these devices. These facts make it possible to create extremely realistic noise maps.

Ubiquity of smartphones enables ordinary citizens to participate in collecting data on noise pollution. Further, the citizens could give feedback to the subjective sound environment and thus further contribute to the system. This way of collecting data is closely linked to the concepts of Citizen science and crowdsensing.

Citizen science represents a potential transformation of models and modes of public participation in science. The term Citizen science is currently used to define a wide range of practices. From citizens who share the processing power of their computers to make scientific calculations possible, through amateur naturalists, residents of cities participating in air pollution measurement, to patients who share symptoms and experiences related to their health problems.

The goal of this paper is to analyze the concepts of Citizen science, and the possibilities to involve citizens in traffic noise measuring using mobile crowdsourcing. As a proof of concept, an experiment has been conducted with the students of Faculty of Organizational Sciences, University of Belgrade, who collected traffic noise data for two weeks. Data was analyzed using statistical methods, in order to find clusters of traffic noise.

LITERATURE REVIEW

Science policy analyst Alan Irwin and ornithologist and participatory research organizer Richard Bonney are often credited with coining the term "citizen science" (Irwin, 1995; Bonney et al., 1996). In Irwin's book Citizen Science: A Study of People, Expertise and Sustainable Development, term Citizen science explains the relationship between citizens and science as a two-way street. Citizen science can be interpreted as a science that serves the interests of the citizens, just as "military science" serves the interests of the military. On the other hand, it is a science performed by citizens. Both interpretations can be explained as science for citizens realized by citizens.

Citizen Science is most often focused on creating and disseminating scientific knowledge beyond scientific institutes but following the norms and values of institutional science.

Richard Bonney, in his 1996 Citizen Science paper, explained this term as a scientific project involving amateurs that provide observational data to scientists, which in turn give them new knowledge based on this data (Booney, 1996; Irwin, 1995). Bonney viewed Citizen science as an opportunity for public participation in scientific research and as a means of promoting science to the public. In 2013, the SOCIENTIZE Expert group for the European Commission's Digital Science Unit defined "citizen science" in a way:" Citizen science refers to the general public engagement in scientific research activities when citizens actively contribute to science either with their intellectual effort or surrounding knowledge or with their tools and resources" (*Green paper on Citizen Science for Europe: Towards a society of empowered citizens and enhanced research | Digital Single Market*, n.d.). The main objective of this approach is to educate the public by letting them participate in scientific research and experiments (Serrano, 2013).

In 2014, the Oxford English Dictionary added its definition of this term: "citizen science: n. scientific work undertaken by members of the general public, often in collaboration with or under the direction of professional scientists and scientific institutions"(*Oxford English Dictionary*, 2014).

Typology of the term

There are a number of types of Citizen science projects because, in most cases, the authors defined the typologies depending on the topic they were addressing. The most commonly used typology of projects is provided by Richard Bonney, who distinguishes:
contributory projects, designed by scientists, where the public contributes primarily to the data;
collaborative projects, where the public can enhance the existing project design, analyze data and disseminate findings;
co-created projects, designed by scientists and members of the public, at least a few of whom are actively involved in most or all steps of the scientific process; (Bonney et al., 2009).

This typology is characterized by a hierarchy, in which co-created projects have more weight because they involve the public or its representatives in higher levels of the process, which is not the case with other types. Later on, this typology was expanded into five modes based on the level of participation:
contractual
contributory
collaborative
co-created
collegial (Shirk et al., 2012).

In this modification, the authors were careful to avoid any hierarchical interpretation, insisting that all of the elements represented a "spectrum."

Alternative typologies are focused, on the goals of the projects as well as the environment in which the projects are executed. Scientists Andrea Wiggins and Kevin Crowston (Wiggins & Crowston, 2011) distinguish five types of Citizen science:
action (reaching local civic agendas through science),
conservation (management of natural resources),
investigation (data collection in a natural environment)
virtual (online scientific research projects),
education (science education in formal and informal settings)

The five activities or practices that need to be implemented during participatory projects are:
sensing,
computing,
analyzing,
self-reporting,
decision making.

These activities help us see beyond the recent initiatives carrying the label "citizen science" and to capture the broader picture of a participatory project. This typology does not imply any hierarchical distinction between the elements, because they are qualitatively different and, most often, by interaction they generate the wanted knowledge. Their purpose is to help us analyze rather than classify participatory projects. Citizen science encompasses a number of methodologies that support significant public contributions to advance scientific and engineering research and monitoring, including:
Identification of research questions
Conducting scientific research
Data collection, processing and analysis
Hardware and software development
Solving complex problems

As one of the methods for realization of Citizen Science, crowdsourcing engages a large group of people through an open call to solve a common task or problem, either individually or collectively. A narrower term than crowdsourcing is crowdsensing. Using crowdsensing techniques, citizens are invited to use smart mobile devices to collect and share information about measurements in order to achieve common interest.

Citizens who participate in Citizen science projects are referred to as Citizen scientists. These are citizens participating in the research who are not formally trained scientists. The number of such citizens has increased significantly in the past decade with the advent of new technologies, the use of mobile phones for data collection and the growth of data sharing, especially in the areas of environmental health. A number of other factors have contributed to the increase in the number of citizen scientists in research, including improvements in accuracy and reduced sensor costs, the benefits of community participation in research, and a reduction in research budgets.

These changes have fostered a new approach to environmental health research that is increasingly reliant on community involvement and which enables the collection of large amounts of data available to the public and greater transparency in decision-making processes. Despite the benefits of using this approach, some researchers have raised concerns about whether the Citizen science approach produces data that can be defended as scientifically accurate and valid (Aceves-Bueno et al., 2017).

Participatory research, in its full form, involves researchers and communities in all aspects of a research project, including the formation of project questions, data collection processes, analysis, interpretation and dissemination of the project. Emphasis is placed on the transition from research to action; researchers and communities work together on actions based on research results and changes in community behavior, laws, and policies to improve the environment.

The range of participatory research approaches is illustrated in Figure 1. Citizen participation ranges from solely collecting data to being fully involved in defining problems and engaging in actions to improve the environment. Volunteered Geographic Information (VGI) or crowdsensing techniques, where a large number of individuals can sign up to provide information, fall to the bottom, along with participants who actively transmit data or simply enable passive data collection.

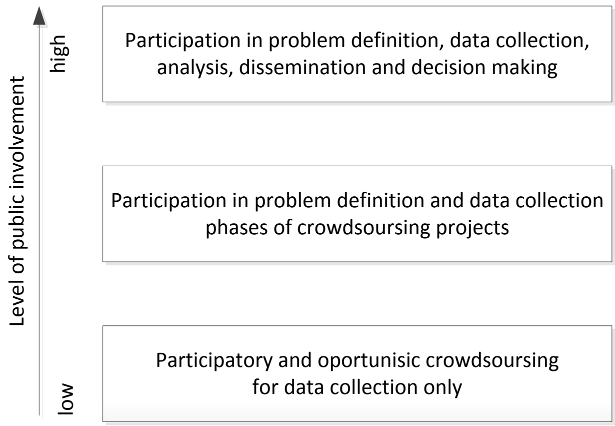

FIGURE 1. The levels of public involvement in participatory research

Barriers to access and literacy in the use of technologies such as computers, the Internet and smartphones can prevent disadvantaged communities from engaging in data collection or transmission projects. However, there are alternative ways of collecting geographic data, which can be applied in communities with limited access or knowledge of information and communication technologies or otherwise influenced by the digital divide. Community members can provide geographic information by drawing or creating their own maps by physically marking the locations of places featured on printed maps.

The centerpiece of the pyramid is a limited participatory research approach that involves the public not only in collecting data but also in defining the problem. "Extreme" approaches to citizen scientists give an even greater role, that is, involve the public in data analysis, interpretation, and ultimately provide opportunities for citizens to engage in actions that will lead to environmental improvements.

Figure 2 shows the ideal scenario for collecting and sharing data in a Citizen science project.

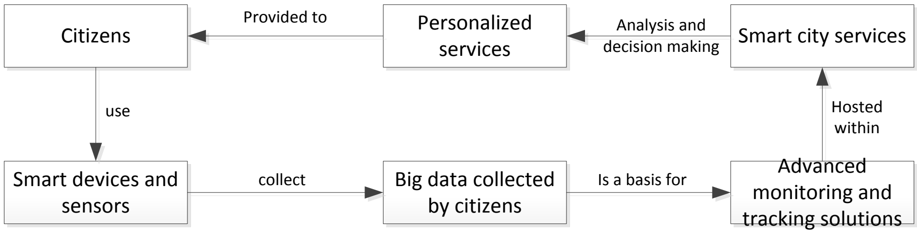

FIGURE 2. The lifecycle of data collection and sharing in "ideal" Citizen science projects

Democratization

The European Commission has unambiguously defined that Citizen science enables the democratization of science (*Citizen science | Digital Single Market*, n.d.). The Zooniverse Crowdsourcing Platform (*Zooniverse - About*, n.d.) defined Citizen science as “people-enabled research” and the Scientific American blog as “people science, from people, to people” (*Science of the People, by the People and for the People - Scientific American Blog Network*, n.d.).

Democracy can be about many things, but if a single element is to be found in all theories of democracy, it is that power must be distributed among all citizens (Christiano, 2015). Something is more democratic if more people are involved. Citizen science stands in contrast to traditional science, as a closed and inaccessible to public. The organizers of the Citizen Science project emphasize that anyone can become a citizen scientist.

In the further text, we present a Citizen science approach to collecting traffic noise data in a city. The presented approach follows the lifecycle presented in figure 2.

EXPERIMENTAL DESIGN

The research included gathering data through crowdsensing noise measurement. Students from the Faculty of Organizational Sciences, University of Belgrade participated in the measurements. To measure the noise, students used a mobile Android application, every day from 8th of May to 28th of May, 2019. The measurement length was 30 seconds and the distance between each measurement was 1 minute. Measurements were made at different microlocations in Belgrade by recording 5 to 10 measurements per microlocation, per day, per student. There were 143 locations measured in total, of which locations Vojvode Stepe, Bulevar Oslobodjenja, Jove Ilica, Studentski grad and Bogoslovija were primary. The collected data was stored in MondoDB database.

Figure 3 shows a map of Belgrade with listed measurement locations. Red circles represent measurement locations where noise is measured over 60 dB, orange circles indicate noise levels between 30 dB and 60 dB, and green noise levels below 30 dB. The data sample included 4251 measurements, ie 109 measurements per day, or 76 measurements in the period 08-10h and 16-18h.

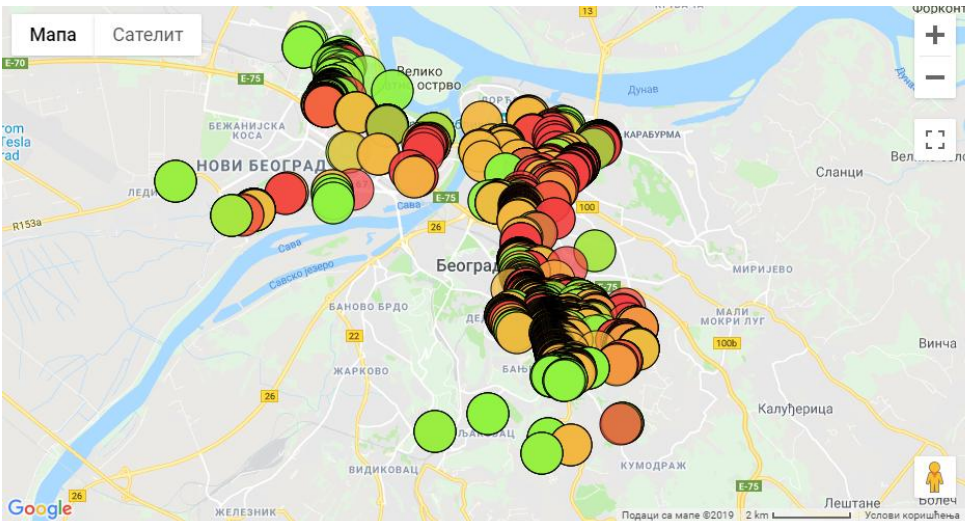

FIGURE 3. Map of Belgrade with listed measurement location

Statistical analysis was performed using Excel, R and R Studio.

RESULTS AND DISCUSSION

We grouped and performed the noise measurements into five locations. The measured values are shown in Table 1. Table shows for each site, the average value of all measurements, the average value of measurements in periods 8-10h and 16-18h, and maximum value mesaured.

TABLE 1. Measurement locations and measured average and maximum values

LOCATION	AVERAGE VALUES	AVERAGE VALUES 8-10h and 16-18h	MAXIMAL VALUES
Bogoslovija (heavy traffic roundabout)	45.3 dB	48 dB	69 dB
Bulevar Oslobodjenja (heavy traffic 4-lane street)	35.4 dB	38 dB	101 dB
Jove Ilica (residential street)	35.5 dB	39 dB	104 dB
Studentski grad (residential area near highway)	40.5 dB	42 dB	73 dB
Vojvode Stepe (heavy traffic 2-lane street)	43.3 dB	49 dB	102 dB

Figure 4 shows average and maximal measured values for one selected location – traffic circle Bogoslovija. Based on the average values, we conclude that the measured noise levels in Belgrade are within normal limits. However, the results obtained represent averages. If we look at the maximum recorded values shown in Table 1, we can conclude that at some moments the measured noise exceeds the level of the prescribed limit values. The noise level Bulevar Oslobodjenja measured in the period 16-17h was 101 dB, which is 36 dB more than the maximum value. Table 1 also shows that at the locations Bogoslovija, Bulevar Oslobodjenja, Jove Ilica and Studentski grad, in the periods 08-10h and 16-18h, when the traffic density is highest, the volume of noise was increased by 2.7 dB, 2.6 dB, 3.5 dB, 1.5 dB respectively.

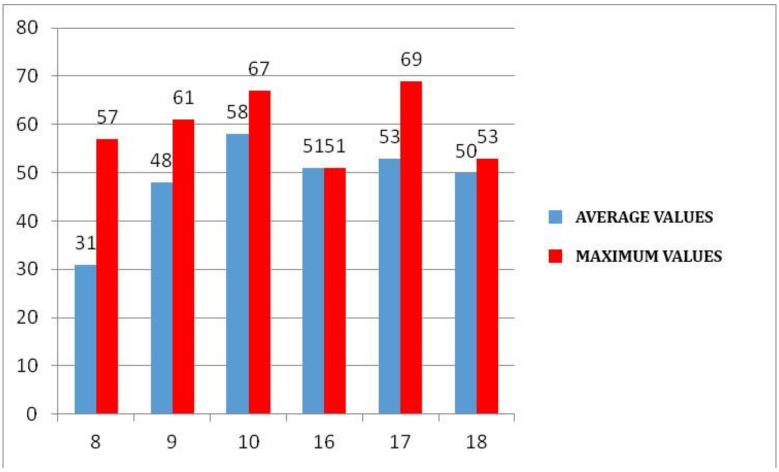

FIGURE 4. Display of measurements of average and maximum values per hour in the periods from 08 to 10h and from 16 to 18h on Bogoslovija location

Noise levels are significantly lower in Jove Ilic Street and in the vicinity of the Studentski grad location, as noise was measured at locations that are residential and not predominantly affected by traffic jams. Higher values were measured at the remaining two locations, since the largest number of measurements was made in the vicinity of the Faculty of Organizational Sciences, University of Belgrade and near the roundabout Bogoslovija, which are among the busiest roads in Belgrade and in addition to passenger cars, buses and trams operate on this location.

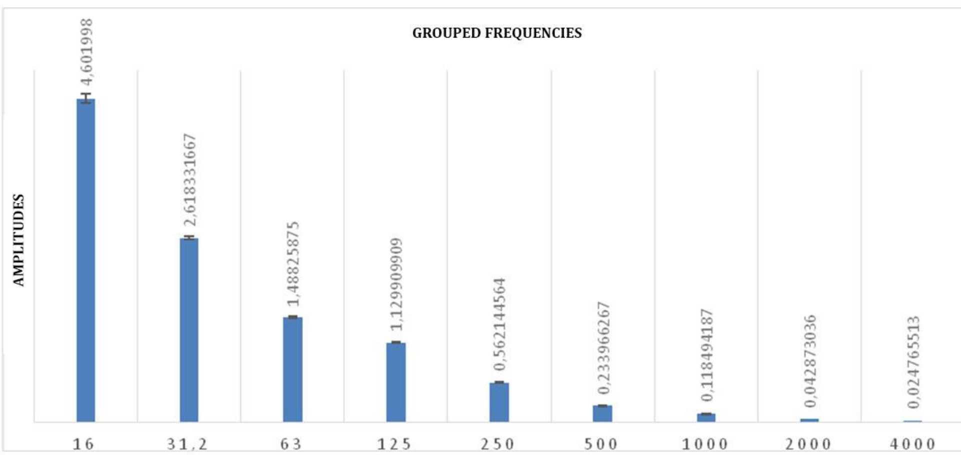

FIGURE 5. Display of the most prominent group frequencies and their amplitudes with deviations

Further analysis was done through considering the spectral distributions of the measured noise. Figure 5 shows that at higher frequencies, the amplitudes tend to zero. The highest amplitude occurs around the frequency of 16 Hz. There is also a maximum deviation, that is, a standard deviation of ~ 0.07 Hz. As frequency increases, amplitudes and standard deviations decrease. In the further analysis, the frequencies from 0 to 100 Hz are considered with more details, since the largest amplitudes and deviations occur at this interval.

Data analysis using statistical methods

The resulting dataset has 650 variables and 4251 observations. Due to the size of the data set and its specific features, as well as for the sake of successful analysis, the dimension of the given dataset was reduced. By examining the frequency values, it has been found that they decrease as the frequency increases. The new dataset has been cleared of "specific values" (outlier) by reducing all values above the ninety-fifth percentile to its value.

The analysis was started by determining the dependence of the selected variables. Since these are quantitative variables, the Pearson correlation was applied (Kovacic, 2004), and the results are presented in Figure 6.

The diagonal elements of the matrix represent frequencies from 0 to 100. Below the diagonal are the correlation coefficients of the frequency pairs, while above the diagonal are graphical representations of these coefficients. To the right of the matrix is a legend. From the presented matrix, it can be concluded that most frequencies are highly correlated with each other. For example, frequencies 5 and 10 are highly positively correlated ($\rho = 0.93$), while they are slightly less correlated with other frequencies, so it might be interesting to examine their relationship in more detail. On the basis of the obtained results, it was decided to apply Principal Component Analysis.

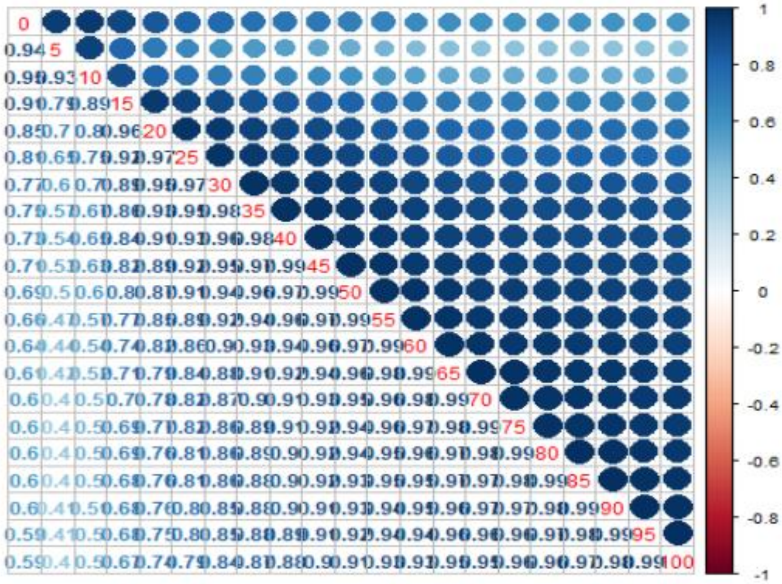

FIGURE 6. Frequency correlation matrix from 0 to 100Hz

Principal component analysis is a multivariate analysis method used to reduce the dimension of a data set made up of a large number of mutually correlated variables, while maintaining the maximum possible variability present in data. The main task of the principal component method is to determine the linear combination of the original variables that will have maximum variance. The second, more general task is to determine several linear combinations of the original variables, which, in addition to having maximum variance, will be uncorrelated among themselves, losing as little as possible the information contained in the set of original variables. In this procedure, the original variables are transformed into new ones, which we call the principal components. The first component is designed to capture the largest part of the variance, while each subsequent component contains the part that is not yet covered (Kovacic, 2004; Pallant et al., 2011). Before approaching the method described above, it is necessary to determine whether the data at its disposal are suitable for its application, and what the appropriate number of principal components is.

KMO and Bartlett's Test[a]

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.961
Bartlett's Test of Sphericity	Approx. Chi-Square	293255.540
	df	210
	Sig.	.000

FIGURE 7. KMO and Bartlett’s test

The Kaiser-Meyer-Olkin coefficient (KMO) (Pallant et al., 2011) shows the "convenience" of applying principal component analysis to the given data and takes the value from the interval [0,1]. A KMO with a value greater than 0.6 is considered adequate. As the KMO in Figure 7 equals 0.961, the principal components analysis is considered justified. Bartlett's test of sphericity also helps in this decision (Pallant et al., 2011). With a p-value of less than 0.01, one can reject the null hypothesis and conclude that the data are not spherical.

The following figure shows the result of the so-called "Scree test", which looks at the eigenvalues of the respective components and then looks for the largest fracture of line that connects them. The components to be selected are those located to the left of the fracture. As in this case the fracture is on component three, the method proposes two components to continue the analysis (Pallant et al., 2011).

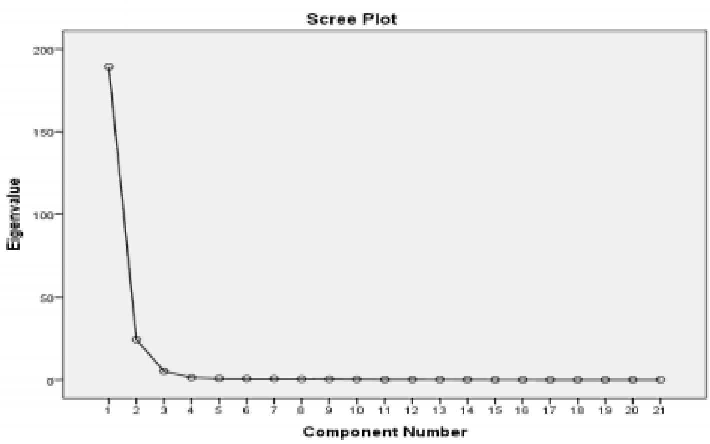

FIGURE 8. Scree Plot, “Elbow” method

The results of the principal component analysis method, obtained by orthogonal Varimax (Kovacic, 2004) rotation are shown in Figure 9.

Rotated Component Matrix

	Raw		Rescaled	
	Component		Component	
	1	2	1	2
@70	2.357	.695	.934	.275
@65	2.403	.760	.933	.295
@75	2.228	.646	.931	.270
@80	2.056	.594	.928	.268
@85	1.963	.571	.924	.269
@60	2.491	.886	.924	.329
@90	1.851	.548	.919	.272
@95	1.820	.539	.914	.271
@55	2.571	1.047	.913	.372
@100	1.799	.527	.906	.265
@50	2.740	1.269	.896	.415
@45	2.969	1.552	.874	.457
@40	3.246	1.864	.853	.490
@35	3.438	2.183	.830	.527
@30	3.386	2.488	.790	.580
@25	3.152	2.876	.719	.656
@10	1.156	4.863	.226	.950
@5	.326	2.551	.118	.921
@0	.450	1.163	.347	.896
@15	2.395	3.934	.510	.838
@20	2.993	3.394	.645	.731

FIGURE 9. Results of principal components analysis after rotation

From the Figure 9, it can be concluded that the frequencies 0, 5, 10, 15 and 20 belong to the component two, while the other frequencies belong to the component one.

As it has been shown that it is appropriate to use such components, further testing will be conducted by analyzing the phenomena and properties of the data only through these two components. This reduced the size of the original dataset, which was the goal. The possibility of multicollinearity (James et al., 2013), which is due to the high correlation of several variables, is also reduced. Its effects adversely affect the analysis, so it is advisable to eliminate them.

Analyzing frequency values through the days of the week

Mean values of lower frequencies (elements of the component two) by different days of the week are the following:

```
                  Mon.      Tue.       Wed.      Thu.       Fri.     Sat.      Sun.

mean  values     27.1791   23.6537   23.9671   19.4233   26.6020   19.7685
24.1258
```

The presented data shows that mean values on Thursdays and Saturdays seem to be lower, comparing to other days of the week. Figure 10 shows a difference in frequency values on Thursdays and Saturdays compared to other days. However, as these differences may also occur "by chance", depending on the sample, the given graph is not sufficient to reach a final conclusion. From Figure 10. it is possible to see differences in the "spread" of lower frequency values within groups (the group in this case represents the day of the week).

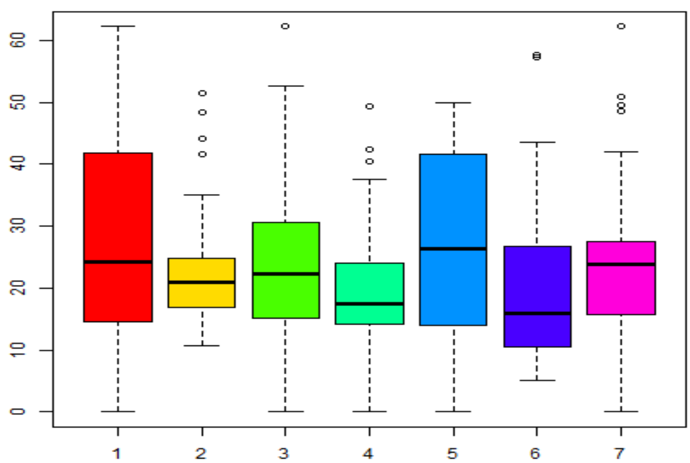

FIGURE 10. Frequencies by days of week

Cluster analysis - Hierarchical clustering

Cluster analysis is a multivariate analysis method that generates groups, such that the within-group objects are similar to each other but at the same time, significantly different from the objects belonging to other groups. There are two basic types of cluster analysis: non-hierarchical and hierarchical clustering (Kovacic, 2004).

Hierarchical methods yield a series of consecutive partitions of a set to clusters, first calculating the distances of all objects to each other, and then forming clusters using merging or splitting techniques. A set of clusters is formed and organized in the form of a tree, which is most often shown using a dendrogram. Hierarchical methods can be classified into association methods or agglomerative methods and divisional or divisive methods, depending on how hierarchical decomposition is formed (Dragic, 2015; James et al., 2013; Kovacic, 2004; Pallant et al., 2011).

Hierarchical agglomerative methods are most commonly used. Starting with n objects with each observation representing a single cluster and sequentially merging into larger clusters (Figure 11).The process ends when all objects belong to the same group (Dragic, 2015; James et al., 2013; Kovacic, 2004; Pallant et al., 2011).

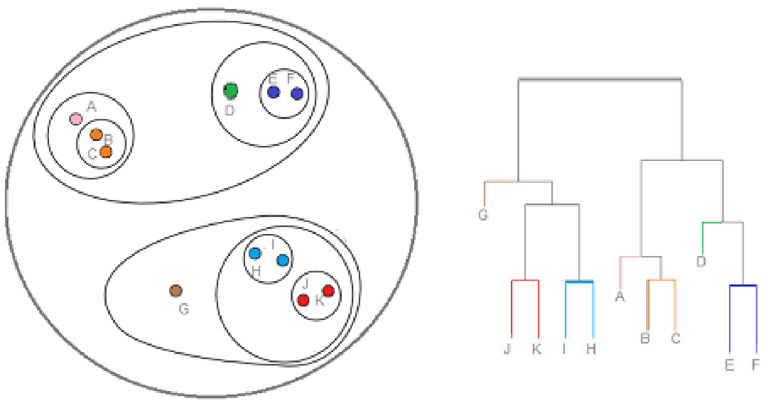

FIGURE 11. Hierarchical clustering (Hierarchical Clustering / Dendrogram: Simple Definition, Examples - Statistics How To, n.d.)

Agglomerative methods differ in how the distance between clusters is estimated in each of the consecutive steps. The most commonly used is Lance-Williams (Dragic, 2015) group of methods, including:
Centroid linkage
Single linkage
Complete linkage
Average linkage

Ward’s method

It was decided to use Ward's method as well as the centroid method in the following work. The aim of this analysis is to group the eleven locations with the highest number of measurements into three groups (highest noise, medium or neutral noise and minimum noise).

Ward’s method

Ward’s Method or Minimum variance method, like other agglomerative methods, starts with n clusters (each cluster contains one observation) but it does not calculate distance between clusters, instead, it maximizes homogenety within the cluster. The total sum of squares within a cluster (SSE) is calculated to determine which two groups merge at each step of the algorithm. SSE is defined:

$$\text{SSE} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - x^i)^2,$$

where x_{ij} is j*th* object in i*th* cluster, k represents the number of clusters, x^i is center of the i*th* cluster and n_i is the number of objects in i*th* cluster. Ward's method does not require the assumption of a multivariate normal distribution of variables (Dragic, 2015; James et al., 2013; Kovacic, 2004; Pallant et al., 2011).

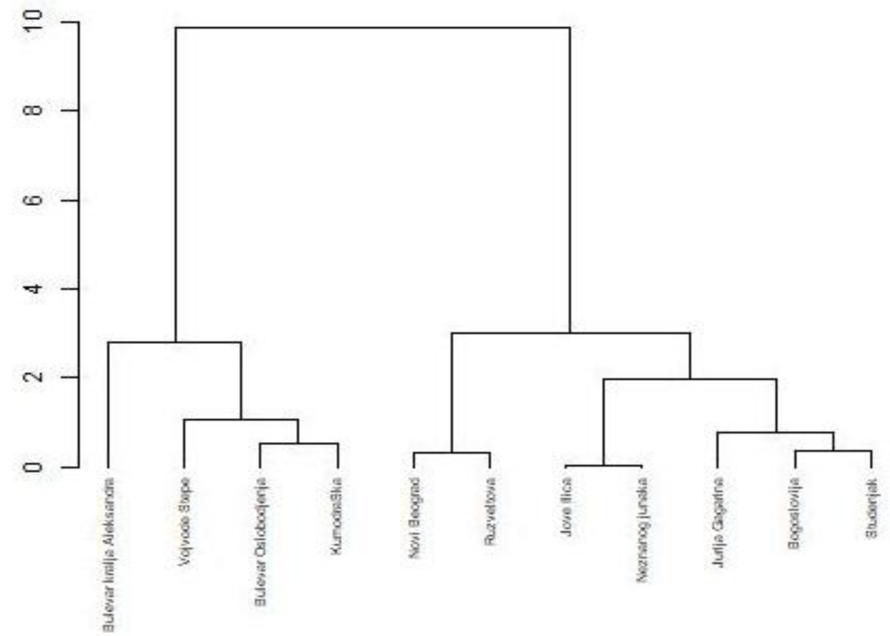

FIGURE 12. Hierarchical clustering – initial dendrogram

The initial dendrogram created using the measured traffic noise data is shown in the Figure 12. From the initial dendrogram it is possible to see the grouping process from the lowest levels to the final group. We generated the following groups from the first iterations: Bulevar Oslobodjenja and Kumodraška, Novi Beograd and Ruzveltova, Jove Ilića and Neznanog junaka, Bogoslovija and Studentski grad. We see that the other locations remain unbundled after the first few iterations. Vojvode Stepe is then assigned to the first group. The process shown continues until all locations are in the same group.

The locations are then divided into three groups:

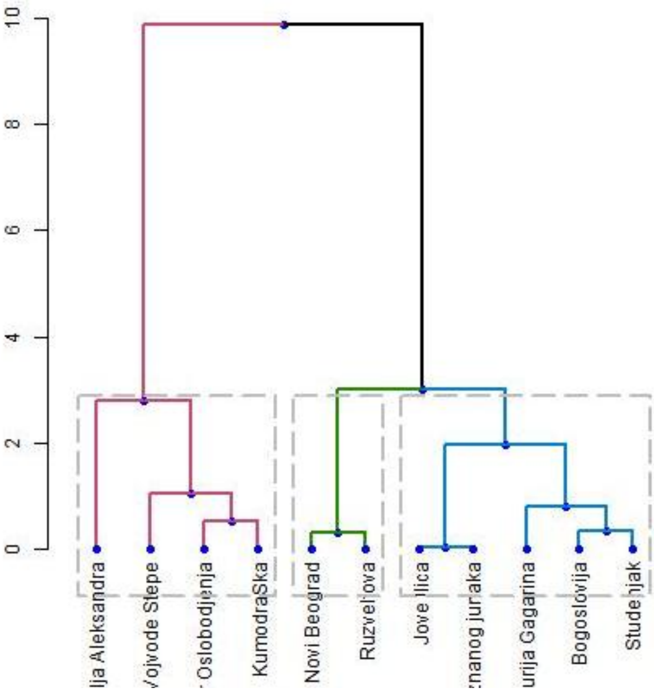

FIGURE 13. Three-group dendrogram – Ward's method

From the shown dendrogram (Figure 13), final groups can be observed. The locations assigned to the first group are: Bulevar kralja Aleksandra, Vojvode Stepe, Bulevar Oslobodjenja and Kumodraska. The locations assigned to the second group are: Novi Beograd and Ruzveltova. The locations assigned to the third group are: Jove Ilica, Neznanog junaka, Jurija Gaganira, Bogoslovija and Studentski grad.

By examining the frequency values for each of the locations, we can conclude that the first group consists of locations with high noise, the second one is with less noisy locations, while the third group contains locations with the lowest noise. The result shows that more information can be obtained from the results if data is analyzed on the level of microlocations.

CENTROID LINKAGE

Cluster distance is represented by centroid distance. Two clusters merge if the distance between their centroids is the shortest relative to the distance of all other pairs of clusters existing at the observed level of association (Dragic, 2015; James et al., 2013; Kovacic, 2004; Pallant et al., 2011).

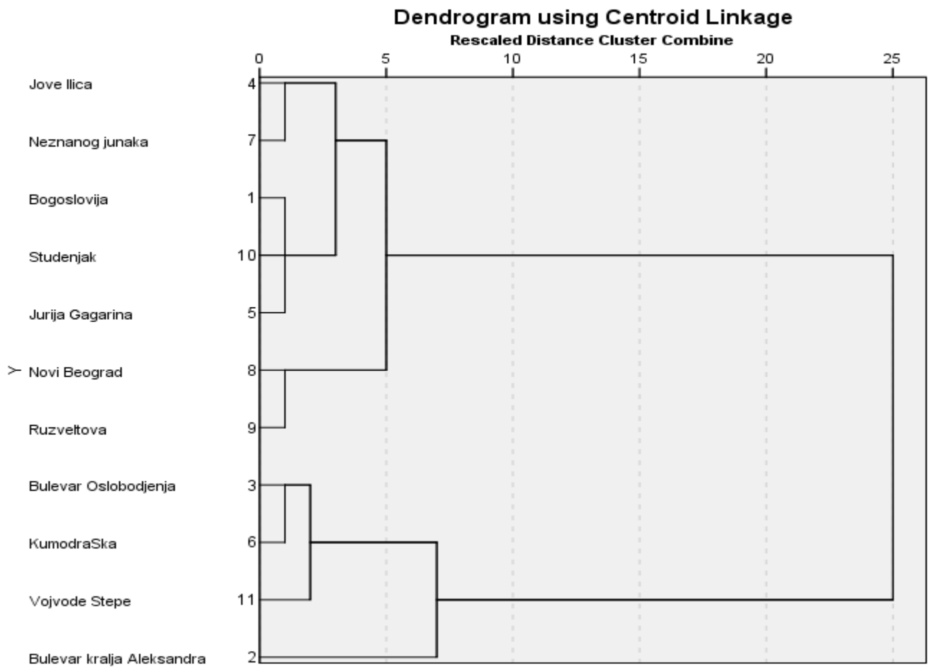

Figure 14. Dendrogram - centroid

Figure 14 shows the dendrogram obtained by using centroid linkage. Based on the given dendrogram, it is possible to conclude the following: Bulevar kralja Aleksandra is location with the highest noise, so it is significantly different from other locations. Locations: Bogoslovija, Jove Ilica, Jurija Gaganira, Neznanog junaka, Novi Beograd, Ruzveltova i Studenjak, belong to the group with the lowest noise. The medium noise group consists of locations: Bulevar Oslobođenja, Kumodraška i Vojvode Stepe.

CONCLUSION

Citizen science can indeed be seen as the next step in a participatory turn, and has the potential to overcome the disadvantages of a deliberative regime by involving citizens in science itself. This approach turns citizens into scientists, promising to produce new knowledge, educating the public and, above all, reconfiguring science from closed to open activity. This movement is "democratizing" science.

With the development of affordable Interet of things devices and mobile phones with high-end functinallities, the citizens become more empowered to collect data through crowdsourcing. In this way they can easily collect data, store and share them using cloud services, provide data to be analyzed by city government or analyze data themselves. The results of analysis can give new insights relateds to microlocations in the city, especially important in case where this type of granularity is not considered by standard measurements, for example measuring the noise level in urban area.

This paper presented a view on citizen sciences in the context of crowdsensing traffic noise data. As a proof of concept, traffic noise data was collected in the city of Belgrade, Serbia. Statistical analysis showes some of the information that can be found using the collected data, but further work is necessary to develop useful decision making tools. The future work will also be oriented toward developing a comprehensive big data infrastructure for storing and real-time analyses of the collected traffic noise data.

REFERENCES

Aceves-Bueno, E., Adeleye, A. S., Feraud, M., Huang, Y., Tao, M., Yang, Y., & Anderson, S. E. (2017). The Accuracy of Citizen Science Data: A Quantitative Review. *The Bulletin of the Ecological Society of America*, *98*(4), 278–290. https://doi.org/10.1002/bes2.1336

Bonney, R., Mccallie, E., & Phillips, T. (2009). *Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education*. www.caise.insci.org

Booney, R. (1996). Citizen Science. *Living Bird*.

Christiano, T. (2015). *Democracy*. *September*, 1–22. https://doi.org/10.1111/1467-9973.00225

Citizen science | Digital Single Market. (n.d.). Retrieved January 5, 2020, from https://ec.europa.eu/digital-single-market/en/citizen-science

Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise - Declaration by the Commission in the Conciliation Committee on the Directive relating to the assessment and management of environmental noise - Public. (n.d.). Retrieved January 5, 2020, from https://op.europa.eu/en/publication-detail/-/publication/27d1a64e-08f0-4665-a258-96f16c7af072/language-en

Dragic, N. (2015). *KONSTRUKCIJA I ANALIZA KLASTER ALGORITMA SA PRIMENOM U DEFINISANJU BIHEJVIORALNIH FAKTORA RIZIKA U POPULACIJI ODRASLOG STANOVNIŠTVA SRBIJE*.

EEA(Europeon Environment Agency), Braubach, M., Jacobs, D. E., Ormandy, D., REC, Umweltbundesamt, Joseph, W. S., Kephalopoulos, S., Paviotti, M., Anfosso-Lédée, F., Hellmuth, T., Classen, T., Kim, R., Kephalopoulos, S., Report, F., Malchaire, J., Braubach, M; Jacobs, D; Ormandy, D., Hammer, M. S., Swinburn, T. K., … European Environmental Agency. (2014). Common Noise Assessment Methods in Europe (CNOSSOS-EU). In *Journal of the American Podiatric Medical Association* (Vol. 122, Issue 2). https://doi.org/10.2788/31776

Green paper on Citizen Science for Europe: Towards a society of empowered citizens and enhanced research | Digital Single Market. (n.d.). Retrieved January 5, 2020, from https://ec.europa.eu/digital-single-market/en/news/green-paper-citizen-science-europe-towards-society-empowered-citizens-and-enhanced-research

Hierarchical Clustering / Dendrogram: Simple Definition, Examples - Statistics How To. (n.d.). Retrieved January 5, 2020, from https://www.statisticshowto.datasciencecentral.com/hierarchical-clustering/

Irwin, A. (1995). Citizen Science -. In *London and New York*. Routledge. https://doi.org/10.1017/CBO9781107415324.004

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. https://doi.org/10.1007/978-1-4614-7138-7

Kovacic, Z. J. (2004). *MULTIVARIJACIONA ANALIZA*.

Oxford English Dictionary. (2014). https://www.oed.com/

Pallant, J., Šućur, M., Milanko, O., & Artprint). (2011). *SPSS : priručnik za preživljavanje : postupni vodič kroz analizu podataka pomoću SPSS-a*. Mikro knjiga.

Science of the People, by the People and for the People - Scientific American Blog Network. (n.d.). Retrieved January 5, 2020, from https://blogs.scientificamerican.com/guest-blog/science-of-the-people-by-the-people-and-for-the-people/

Serrano, F. (2013). *Engaging citizens in science for research excellence*. https://sciencenode.org/feature/engaging-citizens-science-research-excellence.php

Shirk, J. L., Ballard, H. L., Wilderman, C. C., Phillips, T., Wiggins, A., Jordan, R., McCallie, E., Minarchek, M., Lewenstein, B. V., Krasny, M. E., & Bonney, R. (2012). Public participation in scientific research: A framework for deliberate design. *Ecology and Society*, *17*(2). https://doi.org/10.5751/ES-04705-170229

Wiggins, A., & Crowston, K. (2011). From conservation to crowdsourcing: A typology of citizen science. *Proceedings of the Annual Hawaii International Conference on System Sciences*. https://doi.org/10.1109/HICSS.2011.207

Zooniverse - About. (n.d.). Retrieved January 5, 2020, from https://www.zooniverse.org/about