



Revista Facultad de Ingeniería

ISSN: 0121-1129

ISSN: 2357-5328

Universidad Pedagógica y Tecnológica de Colombia

Chanchí-Golondrino, Gabriel-Elías; Ospino-Pinedo, Mónica-Esther; Muñoz-Sanabria, Luis-Freddy
Application of Spatial Data Science on Results of the Saber 5 Test
Revista Facultad de Ingeniería, vol. 30, no. 58, e103, 2021, October-December
Universidad Pedagógica y Tecnológica de Colombia

DOI: <https://doi.org/10.19053/01211129.v30.n58.2021.13823>

Available in: <https://www.redalyc.org/articulo.oa?id=413969810004>

- How to cite
- Complete issue
- More information about this article
- Journal's webpage in redalyc.org

redalyc.org

Scientific Information System Redalyc
Network of Scientific Journals from Latin America and the Caribbean, Spain and Portugal

Project academic non-profit, developed under the open access initiative

Revista Facultad de Ingeniería

Journal Homepage: <https://revistas.uptc.edu.co/index.php/ingenieria>



Application of Spatial Data Science on Results of the Saber 5 Test

Gabriel-Elías Chanchí-Golondrino¹

Mónica-Esther Ospino-Pinedo²

Luis-Freddy Muñoz-Sanabria³

Received: October 11, 2021

Accepted: December 04, 2021

Published: December 10, 2021

Citation: G.-E. Chanchí-Golondrino, M.-E. Ospino-Pinedo, L.-F. Muñoz-Sanabria, "Application of Spatial Data Science on Results of the Saber 5 Test," *Revista Facultad de Ingeniería*, vol. 30 (58), e13823, 2021. <https://doi.org/10.19053/01211129.v30.n58.2021.13823>

Abstract

The purpose of the Saber tests is to improve the quality of education in Colombia through periodic evaluations of the competencies development in basic and secondary education students. Although different data mining studies of the Saber tests' results have been carried out, not many are based on spatial analysis of data in the educational field. This article presents a study based on exploratory and spatial analyses of the 2016 Saber 5 test dataset, which aims at identifying the relationship

¹ Ph. D. Universidad de Cartagena (Cartagena-Bolívar, Colombia). gchanchig@unicartagena.edu.co. ORCID: <https://orcid.org/0000-0002-0257-1988>

² Universidad de Cartagena (Cartagena-Bolívar, Colombia). mospinop@unicartagena.edu.co. ORCID: <https://orcid.org/0000-0003-3357-6316>

³ Ph. D. Universitaria de Popayán (Popayán-Cauca, Colombia). lfreddyys@fup.edu.co. ORCID: <https://orcid.org/0000-0002-8172-0530>

between the results of the test and the geographic or spatial distribution of the data. For the development of this research, four methodological phases were established: data selection and collection, correlation analysis, spatial data analysis, obtaining conclusions. Using the free GeoDa tool, a quantile and regression analysis was performed on the results of the Language, Mathematics, and Natural Sciences areas. Similarly, an analysis based on clustering was performed using the KMeans algorithm. The results obtained made it possible to determine the linear and spatial correlation between the data and the different departments of Colombia at the level of the three areas considered, which may be of interest in decision-making processes by the Ministry of Education. In addition, the research allowed to verify the usefulness of the GeoDa tool for conducting studies based on exploratory and spatial analysis in different application contexts.

Keywords: exploratory analysis; GeoDa; Saber test; spatial analysis.

Aplicación de la ciencia de datos espaciales sobre los resultados de las pruebas Saber 5

Resumen

Las pruebas saber tienen como objetivo contribuir al mejoramiento de la calidad de la educación en Colombia a través de la realización de evaluaciones periódicas que buscan evaluar el desarrollo de competencias en los estudiantes de educación básica. Aunque se han realizado diferentes estudios basados en minería de datos sobre los resultados de las pruebas saber, se han evidenciado pocos estudios basados en análisis espacial de datos en el ámbito educativo. En este artículo se propone como aporte el desarrollo de un estudio basado en análisis exploratorio y análisis espacial sobre el conjunto de datos de las pruebas saber 5 del año 2016, el cual tiene por objetivo identificar la relación entre los resultados de las pruebas saber y la distribución geográfica y/o espacial de los datos. Para el desarrollo de esta investigación se definieron 4 fases metodológicas a saber: selección y obtención de los datos, análisis de correlación, análisis espacial de los datos, obtención de conclusiones sobre el estudio desarrollado. A través del uso de la herramienta libre GeoDa se realizó un análisis de cuantiles y regresión sobre los

resultados obtenidos en las áreas de Lenguaje, Matemáticas y Ciencias Naturales. Del mismo modo, se realizó un análisis basado en clustering mediante el uso del algoritmo KMeans. Los resultados obtenidos permitieron determinar la correlación lineal y espacial existente entre los datos asociados a los diferentes departamentos de Colombia a nivel de las 3 áreas consideradas, los cuales pueden resultar de interés en cuanto a la toma de decisiones por parte del Ministerio de Educación. Así mismo, la investigación desarrollada permitió verificar la utilidad de la herramienta GeoDa para la conducción de estudios basados en análisis exploratorio y espacial en diferentes contextos de aplicación.

Palabras clave: análisis espacial; análisis exploratorio; GeoDa; pruebas Saber.

Aplicação de ciência de dados espaciais em resultados do teste “Saber 5”

Resumo

O objetivo de conhecer provas é contribuir para a melhoria da qualidade da educação na Colômbia por meio de avaliações periódicas que buscam avaliar o desenvolvimento de competências em alunos da educação básica. Embora existam diversos estudos baseados em mineração de dados sobre os resultados dos testes de saber, poucos estudos baseados em análise espacial de dados na área educacional foram evidenciados. Este artigo propõe como contribuição o desenvolvimento de um estudo baseado na análise exploratória e análise espacial sobre o conjunto de dados dos testes de saber 5 do ano de 2016, que visa identificar a relação entre os resultados dos testes de saber e os ou distribuição espacial dos dados. Para o desenvolvimento desta pesquisa, foram definidas 4 fases metodológicas, a saber: seleção e obtenção de dados, análise de correlação, análise espacial dos dados, obtenção de conclusões sobre o estudo desenvolvido. Por meio da ferramenta gratuita GeoDa, foi realizada a análise de quantis e regressão dos resultados obtidos nas áreas de Linguagem, Matemática e Ciências Naturais. Da mesma forma, uma análise baseada em agrupamento foi realizada usando o algoritmo KMeans. Os resultados obtidos permitiram determinar a correlação linear e espacial entre os dados associados aos diferentes departamentos da Colômbia ao nível das 3 áreas consideradas, o que pode ser de

interesse em termos de decisão do Ministério da Educação. Da mesma forma, a pesquisa realizada permitiu verificar a utilidade da ferramenta GeoDa para a realização de estudos baseados em análises exploratórias e espaciais em diferentes contextos de aplicação.

Palavras-chave: análise espacial; análise exploratória; GeoDa; teste de conhecimento.

I. INTRODUCTION

The Saber tests correspond to external standardized evaluations applied by the ICFES (Colombian Institute for the Evaluation of Education), whose objective is to evaluate students' performance according to the basic competencies established by the Colombian Ministry of National Education. These tests evaluate the students' performance at the end of the basic and secondary education levels: Saber 3º and 5º in primary education, Saber 9º at the end of lower secondary education, and Saber 11º at the end of secondary education [1]-[5]. These tests aim to adjust the educational policy and stimulate the quantitative and qualitative growth of the country's educational system regarding coverage or number of beneficiaries and quality [6]-[8].

In the last decade, the exams applied by the Ministry of Education that assess the quality of education in Colombia has allowed the dissemination and access to datasets that can be used in studies aimed at characterizing and explaining the performance of students in different competencies associated with diverse areas of knowledge [9]. However, in Colombia, and the specific case of the open data strategy, few datasets are adequate for spatial analysis in the educational context. Most of the studies have focused on non-spatial exploratory analysis and data mining strategies [10]-[16]. The spatial analysis of data involves a set of techniques and models that specifically use the spatial or geospatial reference of each instance belonging to the dataset to determine spatial correlations [17]-[19].

This paper presents an exploratory study based on spatial analysis of the dataset that includes the results of the 2016 Saber 5 test for the areas of Mathematics, Language, and Natural Sciences. This dataset classified the number of students associated with the levels "Insufficient", "Minimum", "Satisfactory", and "Advanced" according to the 33 departments of Colombia and the knowledge areas mentioned above. The study aimed to identify statistical and geospatial relationships in the selected dataset using the free software tool (GPL licensed) GeoDa [20] to obtain value-added conclusions useful for the decision-making processes by the Ministry of National Education and relevant authorities. Likewise, this study is intended to

serve as a reference to be extrapolated to other educational contexts, such as the Saber tests implemented in secondary and higher education.

The rest of the article is organized as follows: Section 2 presents the methodological phases that guided the development of this research, section 3 describes the results obtained from the spatial analysis of the dataset corresponding to the 2016 Saber 5 test, section 4 presents the conclusions and future work derived from the study.

II. METHODOLOGY

For the development of this research, four methodological phases were established: selection of the spatial analysis dataset, correlation analysis, spatial analysis of the data, generation of conclusions from the study (Figure 1).

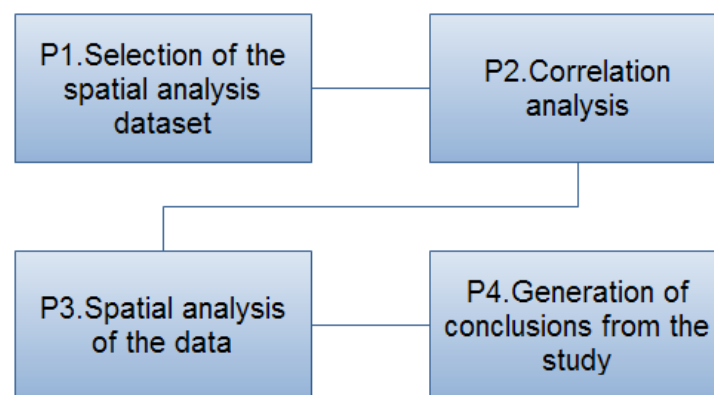


Fig. 1. Methodology considered.

In phase 1 of the methodology, a set of datasets compatible with the spatial data analysis techniques and associated with the Saber test were explored. The datasets available on Colombia's ESRI open data portal were reviewed, and the 2016 Saber 5 test dataset was selected (<https://datosabiertos.esri.co/datasets/esri-colombia::estudiantes-por-nivel-de-desempeno-pruebas-saber-grado-5/about>), which was available in shapefile format (compatible with the GeoDa tool). The selected dataset has a total of 33 records or instances associated with the number of students per department in the "Insufficient", "Minimum", "Satisfactory", and "Advanced" levels in the areas of Language, Mathematics, and Natural Sciences. From the dataset selected in phase 1, a correlation analysis was implemented in

phase 2 on the attributes associated with the student performance levels for the three areas considered. In phase 3, the spatial analysis of the data was carried out by obtaining spatial quantiles and applying clustering models for the "Advanced" and "Insufficient" levels in the areas of Language, Mathematics, and Natural Sciences for the different departments of the country. Finally, in phase 4, the results were consolidated, and the conclusions derived from the study were obtained.

III. RESULTS

The dataset selected for this study includes geospatial data related to the performance of primary school students of the different departments of Colombia in the areas of Mathematics, Language, and Natural Sciences. This dataset has a total of 33 records or rows, which include the number of students per department that are classified in different performance levels ("Insufficient", "Minimal", "Satisfactory", and "Advanced") for the areas mentioned. Thus, Table 1 describes the main attributes associated with the 33 records of the dataset considered in this study.

Table 1. Description of the main attributes of the dataset.

Attributes	Description
DANE	Corresponds to the identifier assigned to each department by the National Administrative Department of Statistics.
Departament	Corresponds to the name of the department of Colombia.
Language_I, Language_M, Language_S, Language_A	Corresponds to the number of students per department whose Saber test score in the area of Language was classified in the following levels: "Insufficient", "Minimal", "Satisfactory", and "Advanced".
Math_I, Math_M, Math_S, Math_A	Corresponds to the number of students per department whose Saber test score in the area of Mathematics was classified in the following levels: "Insufficient", "Minimal", "Satisfactory", and "Advanced".
NScienc_I, NScienc_M, NScienc_S, NScienc_A	Corresponds to the number of students per department whose Saber test score in the area of Natural Sciences was classified in the following levels: "Insufficient", "Minimal", "Satisfactory", and "Advanced".

A. Regression Analysis

This section presents the linear regression analysis of the attributes associated with the "Insufficient" and "Advanced" levels in the areas of Language, Mathematics, and

Natural Sciences. Figure 2 shows the regression analysis of the "Insufficient" level, which allowed to determine a high linear correlation between the pairs of attributes: "Language_I" and "Math_I", "Language_I" and "Math_I", "NScienc_I". The highest correlations correspond to the pairs of attributes: "Language_I" and "NScienc_I", "Math_I" and "N_Scienc_I", with a determination coefficient of 0.974.

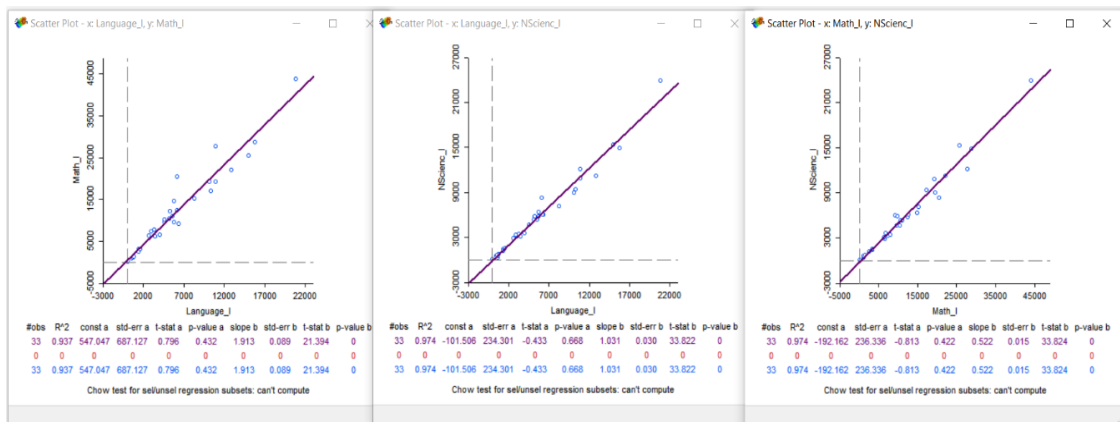


Fig. 2. Regression analysis for the "Insufficient" level.

Similarly, Figure 3 shows the regression analysis of the "Advanced" level of the areas of Language, Mathematics, and Natural Sciences, which allowed to determine a high correlation between the pairs of attributes: "Language_A" and "Math_A", "Language_A" and "NScienc_A", "Math_A" and "NScienc_A". The highest correlation corresponds to the pair: "Language_A" and "NScienc_A", with a determination coefficient of 0.998.

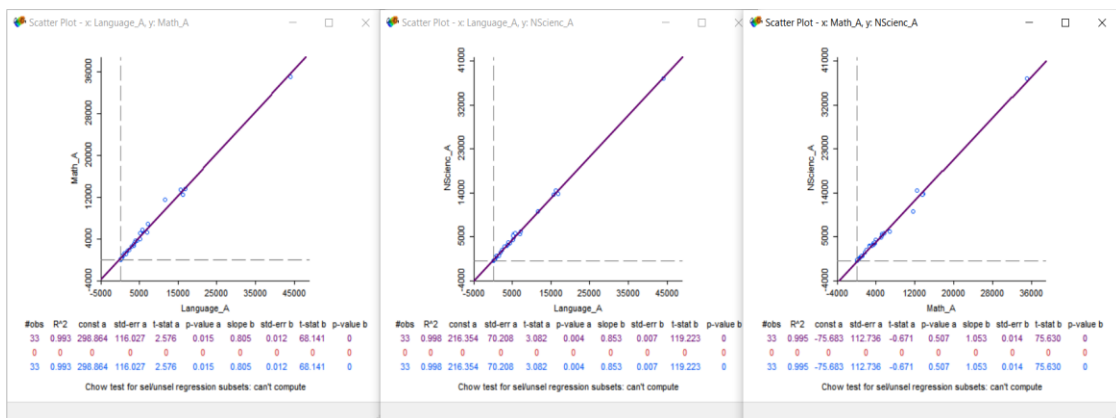


Fig. 3. Regression analysis of the "Advanced" level.

Through the above results, it is possible to conclude that although the linear correlations of the "Insufficient" and "Advanced" levels have a high coefficient of determination, the correlations of the "Advanced" level for the areas of Language, Mathematics, and Natural Sciences are higher than those of the "Insufficient" level.

B. Quantile Analysis

Once the regression analysis was completed, the analysis at the quantile level was implemented in the dataset's attributes. For each area (Language, Mathematics, and Natural Sciences), the geographic distribution of the Advanced and Insufficient levels was considered at 4 quantiles. Thus, Figure 4 shows the 4-quantile analysis for the "Insufficient" and "Advanced" levels in the area of Language (figures on the left and right, respectively). It also shows how the departments of Valle, Antioquia, and Atlántico were grouped for the area of Language in the quantile with the highest number of students for the "Insufficient" and "Advanced" levels (red quantile). Likewise, it is possible to observe how the departments that obtained a greater number of students at the "Advanced" level were: Bogotá D.C., Valle del Cauca, Antioquia, Santander, Cundinamarca, Nariño, Atlántico, and Boyacá. It is also important to mention that the departments of Bogotá D.C., Valle del Cauca, Antioquia, Cundinamarca, and Santander are within the top 5 departments with the best GDPs in the country for 2020 (https://www.dane.gov.co/files/investigaciones/pib/departamentales/B_2015/Bol_PI_B_dptal_2020preliminar.pdf).

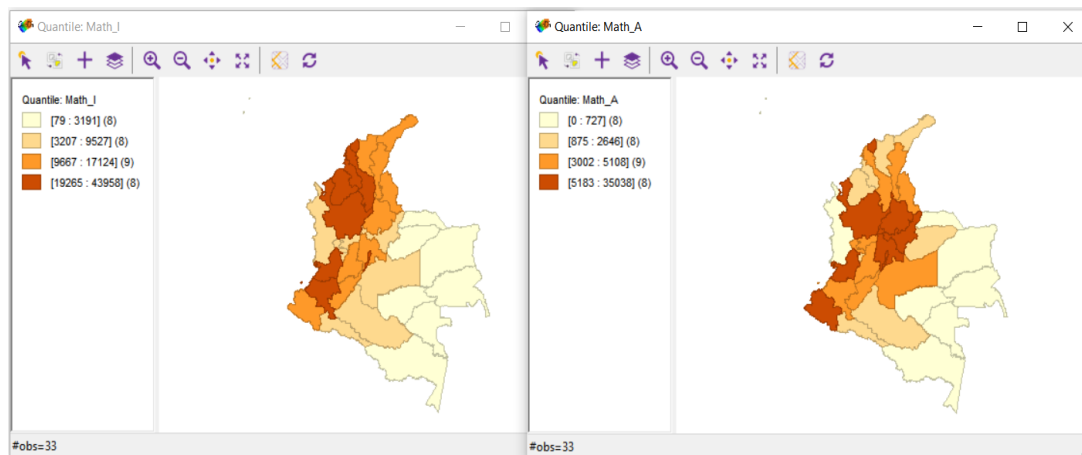


Fig. 4. Four-quantile analysis for the Language area.

On the other hand, Figure 5 shows the 4-quantile analysis for the "Insufficient and Advanced" levels in the area of Mathematics (figures on the left and right, respectively). Additionally, it shows that the departments with the highest number of students at the "Advanced" level of performance were: Nariño, Valle, Antioquia, Cundinamarca, Santander, Bogotá D.C., Boyacá, and Atlántico. As indicated in the analysis of the Language area, 5 of the 8 departments mentioned are in the top 5 (Bogotá D.C., Valle del Cauca, Antioquia, Cundinamarca, and Santander) with the best GDPs in the country for 2020. Consistent with the above, the departments with a lower number of students at the "Advanced" level were the following: Amazonas, Guaviare, Vaupés, Guainía, Vichada, Arauca, Chocó, and San Andrés y Providencia, 6 of which had the worst GDPs for 2020 (Guaviare, Amazonas, Vichada, Guainía, Vaupés, and San Andrés y Providencia).

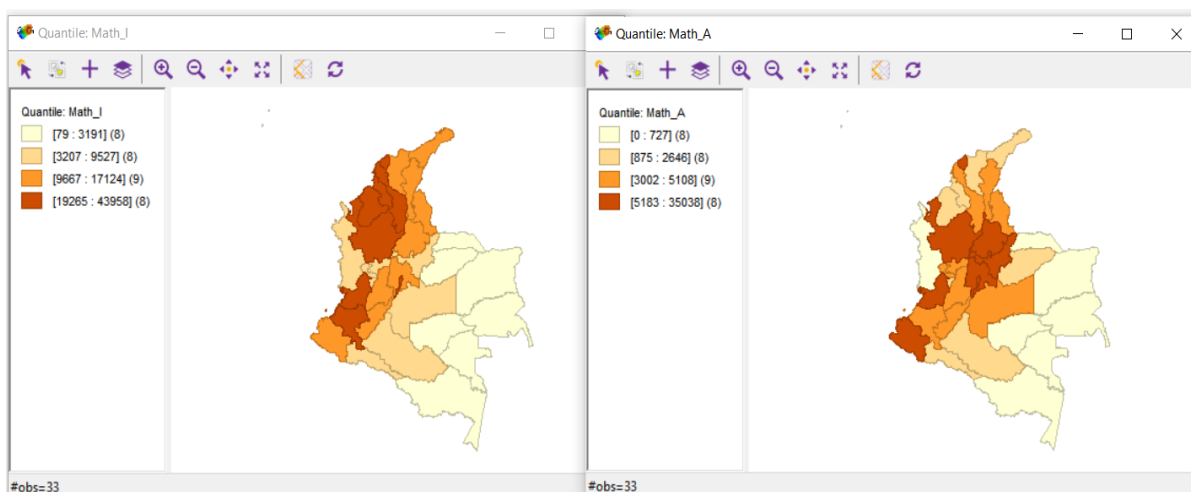


Fig. 5. Four-quantile analysis for the Mathematics area.

Figure 6 shows the 4-quantile analysis for the "Insufficient" and "Advanced" levels in the area of Natural Sciences (figures on the left and right, respectively). It is possible to see from Figure 6 how the departments of Cauca, Valle, and Antioquia were grouped for the area of Natural Sciences in the quantile with the highest number of students for the "Insufficient" and "Advanced" levels (red quantile). Likewise, the departments with the highest number of students at the "Advanced" level were: Nariño, Cauca, Valle, Antioquia, Cundinamarca, Santander, Bogotá D.C., and Boyacá. As indicated in the analysis of the Mathematics area, 5 of the 8 departments mentioned are in the top 5 departments (Bogotá D.C, Valle del Cauca, Antioquia, Cundinamarca, and Santander) with the best GDPs in the country for 2020. In addition, the departments with a lower number of students at the "Advanced" level were: Amazonas, Guaviare, Vaupés, Guainía, Vichada, Arauca, Chocó, and San Andrés y Providencia, 6 of which had the worst GDPs for 2020 (Guaviare, Amazonas, Vichada, Guanía, Vaupés, and San Andrés y Providencia).

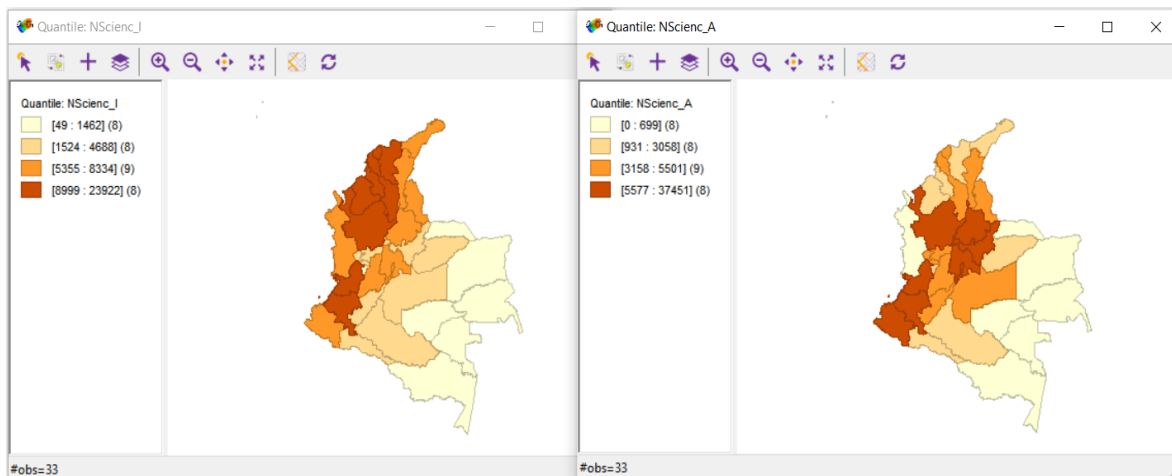


Fig. 6. Four-quantile analysis for the area of Natural Sciences.

C. Clustering Analysis

This section presents the clustering analysis performed for the "Insufficient" and "Advanced" levels in the areas of Language, Mathematics, and Natural Sciences. Thus, Figure 7 shows the clustering analysis after applying the KMeans algorithm with 5 clusters for the "Insufficient" level in the areas of Language, Mathematics, and Natural Sciences.

It is possible to observe from Figure 7 that the cluster including the highest number of students for the 3 considered areas of the test at the "Insufficient" level is cluster 5, whose centroid is $C5=\{\text{Language_I}=20812, \text{Math_I}=43958, \text{NScienc_I}=23922\}$, which geographically corresponds to the department of Antioquia. Likewise, the cluster that groups the lowest number of students for the 3 areas of the test at the "Insufficient" level is cluster 2, whose centroid is $C2=\{\text{Language_I}=821.667, \text{Math_I}=1626.78, \text{NScienc_I}=790\}$, which geographically includes the departments of Amazonas, Vaupés, Guaviare, Guanía, Vichada, Casanare, Arauca, San Andrés y Providencia, and Quindío.

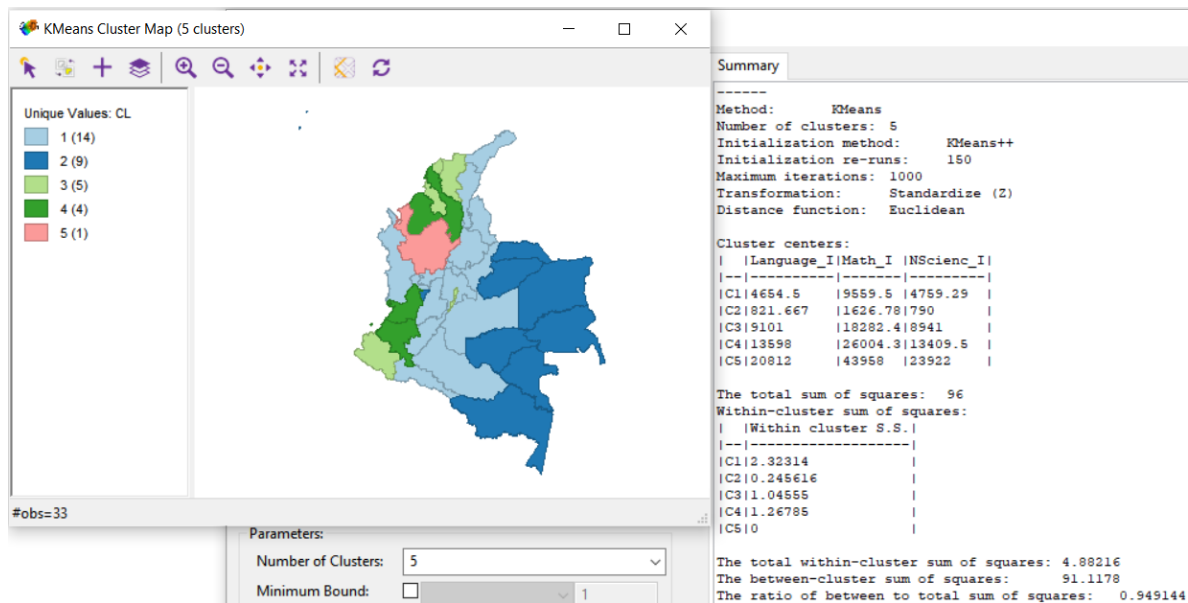


Fig. 7. Clustering analysis of the "Insufficient" level.

Figure 8 shows the clustering analysis using the KMeans algorithm with 5 clusters for the "Advanced" level in the areas of Language, Mathematics, and Natural Sciences.

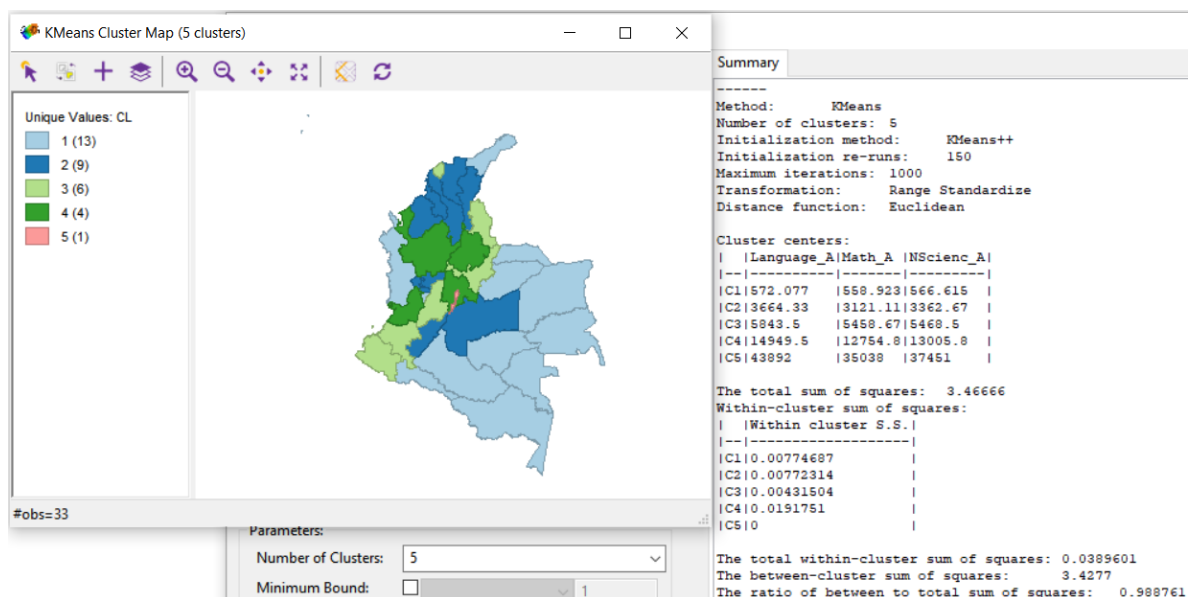


Fig. 8. Clustering analysis of the "Advanced" level.

Figure 8 evidences that the cluster with the highest number of students for the 3 areas of the test at the "Advanced" level is cluster 5, whose centroid is $C5=\{\text{Language_A}=43892, \text{Math_A}=35038, \text{NScienc_A}=37451\}$, which geographically includes Bogotá D.C. Likewise, the cluster that groups the lowest number of students for the 3 areas of the test at the "Advanced" level is cluster 1, whose centroid is $C1=\{\text{Language_A}=572.077, \text{Math_A}=558.923, \text{NScienc_A}=566.615\}$, which geographically includes the departments of: Amazonas, Putumayo, Caquetá, Guaviare, Vaupés, Guainía, Vichada, Casanare, Arauca, Chocó, La Guajira, Quindío, and San Andrés y Providencia.

IV. CONCLUSIONS

This paper presented a study based on spatial data analysis of the dataset with the results of the 2016 Saber 5 test for the different departments of Colombia in the areas of Mathematics, Language, and Natural Sciences. The study included a correlation analysis, a quantile analysis and a clustering analysis using the advantages provided by the free software tool GeoDa. The study developed in this article is intended to be extrapolated to other educational contexts such as secondary and higher education for decision-making processes by the relevant authorities.

The correlation or regression analysis performed on the selected dataset showed a direct and linear relationship between the number of students from the different departments classified in the "Insufficient" and "Advanced" levels for the areas of Natural Sciences and Language. Similarly, through the analysis based on quantiles and clustering, it was possible to determine a direct relationship between the departments with the best GDPs and those with the best performance in the three areas of knowledge considered in the study.

This research allowed to verify the usefulness of the free software tool GeoDa for conducting studies based on exploratory and spatial analysis of data by using linear correlation methods, determination of quantiles, clustering analysis, spatial correlation based on the determination of the Moran index, among others. In this sense, the GeoDa tool is very useful to enrich the exploratory studies carried out in

the educational context, thus, helping in the decision-making processes by the Ministry of National Education and the relevant authorities.

As future research possibilities, there is the application of machine learning techniques based on supervised learning models such as decision trees or association rules after adapting the attributes of the dataset in categorical values.

ACKNOWLEDGMENTS

The authors thank Universidad de Cartagena and Fundación Universitaria de Popayán for the support provided in the development of this project.

AUTHORS' CONTRIBUTION

Gabriel-Elías Chanchí-Golondrino: Supervision, Formal Analysis, Investigation, Writing – Review and editing.

Mónica-Esther Ospino-Pinedo: Conceptualization, Methodology, Writing – Review and editing.

Luis-Freddy Muñoz-Sanabria: Conceptualization, Methodology, Writing – Review and editing.

REFERENCES

- [1] P. Burbano, "Pruebas saber 11: el baremo de la desigualdad educativa en Colombia," *Holopraxis Ciencia, Tecnología e Innovación*, vol. 5, no. 1, pp. 91–108, 2021
- [2] S. M. Chica Gómez, D. M. Galvis Gutiérrez, A. Ramírez Hassan, "Determinantes del rendimiento académico en Colombia. Pruebas ICFES-Saber 11o, 2009," *Revista Universidad EAFIT*, vol. 46, no. 160, pp. 48–72, 2010
- [3] Ministerio de Educación Nacional, *Pruebas saber*, 2020. <https://www.mineducacion.gov.co/1759/w3-article-397384.html?noredirect=1>
- [4] A. Duque Castillo, J. G. Ortiz Rodríguez, "Pruebas ICFES Saber 11 y su relación con el desempeño académico en estudiantes de primer semestre de psicología," *Cuadernos Hispanoamericanos de Psicología*, vol. 13, no. 1, pp. 26–35, 2013. <https://doi.org/10.18270/chps.v13i1.1355>
- [5] O. Vásquez Arrieta, "Las pruebas SABER 11° como predictor del rendimiento académico expresado en los resultados de la prueba SABER PRO obtenidos por las estudiantes de la Licenciatura en Pedagogía Infantil de la Corporación Universitaria Rafael Núñez," *Revista Científica Virtual Hexágono Pedagógico*, vol. 9, no. 1, pp. 187–204, 2018
- [6] M. Castro Ávila, J. Ruiz Linares, F. Guzmán Patiño, "Cruce de las pruebas nacionales Saber 11 y Saber Pro en Antioquia, Colombia: una aproximación desde la regresión geográficamente ponderada (GWR),"

- Revista Colombiana de Educación*, no. 74, pp. 63–79, 2018. <https://doi.org/10.17227/rce.num74-6898>
- [7] W. Martínez-Mateus, Á. Turriago-Hoyos, “Análisis de distribución geográfica y espacial de los resultados de las Pruebas Saber 11 del Instituto Colombiano para el Fomento de la Educación Superior (ICFES). 2005-2012,” *Cuadernos Latinoamericanos de Administración*, vol. 12, no. 21, pp. 39–49, 2015
 - [8] J. E. Cifuentes Medina, J. A. Chacón Benavides, I. Moreno Pinzon, “Análisis de los resultados de las pruebas estandarizadas Saber Pro en profesionales de la educación,” *Revista Humanismo y Sociedad*, vol. 6, no. 2, pp. 22–48, 2018. <https://doi.org/10.22209/rhs.v6n2a02>
 - [9] G. A. Junca Rodríguez, “Desempeño académico en las pruebas Saber 11,” *Panorama Económico*, vol. 27, no. 1, pp. 8–38, 2019. <https://doi.org/10.32997/2463-0470-vol.27-num.1-2019-2615>
 - [10] A. I. Oviedo Carrascal, J. Jiménez Giraldo, “Minería de datos educativos: Análisis del desempeño de estudiantes de ingeniería en las pruebas SABER-PRO,” *Revista Politécnica*, vol. 15, no. 29, pp. 128–140, 2019. <https://doi.org/10.33571/rpolitec.v15n29a10>
 - [11] R. Timarán Pereira, J. Caicedo Zambrano, A. Hidalgo, *Aplicación de la Minería de Datos en la Detección de Patrones de Desempeño Académico en las Pruebas Saber Pro*, Universidad de Nariño, Pasto, Colombia, 2019
 - [12] D. Solís, D. Alegría, É. Gutierrez, V. Zapata, F. Vidal, R. Timarán, “Identificación de Patrones de Rendimiento Académico en las Pruebas Saber Pro entre 2012-2014, en las Competencias Lectura Crítica y Comunicación Escrita con Técnicas Predictivas de Minería de Datos,” *Cuaderno Activa*, vol. 11, no. 1, pp. 51–66, 2019
 - [13] J. S. Rengifo Collazos, C. D. Sánchez Cobo, C. Delgado Manquillo, C. Solarte Sarria, F. Vidal, R. Timarán Pereira, “Análítica de datos aplicada al contexto universitario. Caso de estudio: pruebas Saber Pro,” *Cuaderno Activa*, vol. 12, no. 1, pp. 13–19, 2020
 - [14] R. Timarán-Pereira, A. Hidalgo-Troya, J. Caicedo-Zambrano, “Factores asociados al desempeño académico en Lectura Crítica en las pruebas Saber 11 con árboles de decisión,” *Revista Investigación e Innovación en Ingeniería*, vol. 8, no. 3, pp. 29–37, 2021. <https://doi.org/10.17081/invinno.8.3.4701>
 - [15] R. Timarán-Pereira, J. Caicedo-Zambrano, A. Hidalgo-Troya, “Árboles de decisión para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas Saber 11,” *Revista de Investigación Desarrollo e Innovación*, vol. 9, no. 2, pp. 363–378, 2019. <https://doi.org/10.19053/20278306.v9.n2.2019.9184>
 - [16] L. M. Gómez Melo, Y. A. Jaramillo Córdoba, *Descubrimiento de factores asociados al desempeño en las pruebas saber 5 con técnicas descriptivas de minería de datos*, Grade Thesis, Universidad Tecnológica de Pereira, Colombia, 2017. <http://repositorio.utp.edu.co/dspace/handle/11059/8745>
 - [17] M. Rivero Sánchez, “Análisis espacial de datos y Turismo: Nuevas técnicas para el análisis turístico,” *Revista de Estudios Empresariales*, vol. 2, pp. 48–66, 2008
 - [18] M. Goodchild, R. Haining, “SIG y análisis espacial de datos: perspectivas convergentes,” *Investigaciones Regionales*, no. 6, pp. 175–201, 2005
 - [19] L. L. Sánchez-Peña, “Alcances y límites de los métodos de análisis espacial para el estudio de la pobreza urbana,” *Papeles de Población*, vol. 18, no. 72, pp. 147–179, 2012
 - [20] M. F. Cardenas, J. F. Escobar, K. Gutiérrez, “Equidad territorial en Medellín: espacio público, amenazas naturales y calidad del aire,” *Estudios Socioterritoriales*, no. 27, pp. 1–18, 2020