S3E2: a web-based GIS for the visualization and analysis of socioeconomic segregation in Chile's elementary education system

Julio Rojas-Mora[1], Ignacio Lincolao-Venegas[2], and Felipe Schneeberger-Leon[2]

[1] Dept. of CS Engineering, Universidad Católica de Temuco, Temuco, Chile. `jrojas@inf.uct.cl`
https://www.s3e2.cl/
[2] CS Engineering, Universidad Católica de Temuco, Temuco, Chile.
`[ilincolao,fschneeberger]@inf.uct.cl` https://www.s3e2.cl/

Abstract

S3E2 is a web-based geographic information system (GIS) designed for the visualization and analysis of the socioeconomic segregation of Chile's elementary education system of Chile. It consists of a frontend developed in JavaScript using ReactJS, React-Redux, Leaflet and D3.js, an API developed in Go and ECHO, and a documentary database managed with MongoDB. Data comes from Chile's Ministry of Education, while the provisions of Law 20.248 serve as indicators of vulnerability. S3E2 graphically shows different segregation indices found in the literature, calculated at the commune level. It also allows visualizing, at this same level, the educational institutions that compose it, their basic information, and time series associated with them. S3E2 is a flexible and fast web-based GIS, with a low cost of implementation, due to the usage of free software -or at least free licensing- tools, thus serving as a template for new web-based GIS in different contexts.

1 Introduction

1.1 Problem

After 25 years of democratic governments in Chile, arises the need to reform the education system established in 1981 –during the military dictatorship. The main objectives of this system were, according to Delannoy [12] and Elacqua and Fabrega [14] 1) the decentralization of the administration of public institutions, leaving it in the hands of the municipalities (the so-called "municipalization"), although with a centralized funding source; 2) the promotion of private investment in the education sector and competition between educational institutions, generating a real educational market; and 3) the establishment of the so-called *voucher* system, which allocated funds to schools –regardless of whether they were public or private– to cover tuition for each student who attended them.

In an attempt to improve their ability to achieve vertically upward social mobility [33], advanced students within vulnerable socioeconomic strata migrated to the state-subsidized private education system; this behavior was encouraged by segregating strategies –admission decisions,

expelling failing students, and grouping students according to their academic performance– set up by subsidized schools [5]. Thus, the so-called "peer effect" –the positive externalities or spillover effects that peers of students with high academic performance receive from their social relationship– losses its power in this environment, condemning schools with low academic performance to not have outstanding students that set the example. On the other hand, given that the 1981 education reform has traditionally been classified as "market-oriented", it is to be expected that, as the parents of a high socioeconomic status –highly correlated with high academic performance [27]– can pay more than those of a low one –highly correlated with low academic performance– for the privilege of sending their children to study with those of other parents of a high socioeconomic status, segregation increases; in microeconomic terms, we could say that in an educational system that maximizes profit, segregation will be the result of equilibrium [1].

It is possible to observe confirmation of this theoretical result in empirical studies conducted in the United States [6, 24, 22], where it has been proven that students who attend schools predominantly composed of vulnerable groups tend to have both a lower educational level and, subsequently, a lower remuneration. On the other side of the coin, in [21, 10] it is shown that students belonging to families in vulnerable situations who attend less segregated schools tend to have better academic results.

Given this problematic situation, the objective of this work is the implementation of S3E2 [1], a web-based geographic information system (GIS) that allows for the visualization of the socioeconomic segregation of Chilean basic education.

1.2   State of Art

GIS to support decision-making began to develop as the Web evolved [26], allowing for straightforward visualization of spatial problems, graphically representing different characteristics, whether social, economic and territorial, of a region [23].

From a more specific perspective of development architecture, we can find references for GIS in [18] where they present the development of these systems using free and open-source graphic information systems.

Additionally, in [8], there is a description on modern architectures for GIS, with programming languages like Java with its Spring framework for the process layer, AngularJS for the frontend and libraries such as as Leaflet to generate interaction with the maps. In this work, we can appreciate that with a variety of different tools, it is possible to generate a highly modular display system that facilitates reuse as well as system maintenance.

2   Architecture

Most academic approaches to geographic studies use widely available commercial GIS (such as the ubiquitous ESRI ArcGIS). In this work, an application based on free and open technologies was developed, allowing flexibility and speed when presenting the required information while offering an inherent cost reduction due to not only being *libre* but also *free*.
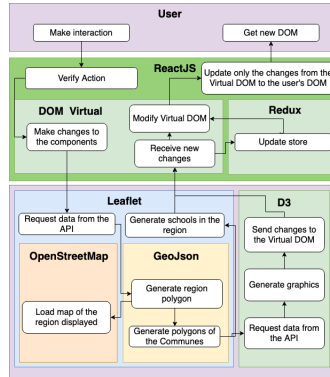
Figure 1: Frontend Architecture.

2.1   Frontend

The frontend of the web-based GIS is structured, as can be seen in Figure 1, in three layers: the user *Document Object Model* (DOM) level, the virtual DOM level and the display component level.

ReactJS [16] works by creating both a user and a virtual DOM; The virtual DOM allows users to reduce the loading time of the DOM tree when created in the user's cache while accessing the platform for the first time. When an interaction with the interface occurs, a change is made in the Virtual DOM, and only the elements that change in the Virtual DOM will be modified in the user's DOM.

ReactJS components form nodes in a hierarchical structure; however, as the application grows, communication between the components would be more efficient if it could be carried out directly instead of going through a long and complicated tree. Hence, our platform uses React-Redux [34], a JavaScript library that centralizes the state of the components of the application through a single tree called "store," allowing for the direct modification of a state when a component is the subject of an action.

The display components are four. For the storage of georeferenced objects, we use GeoJSON files [9], a standard that allows extending the JSON format for the representation of geographical features and their properties. To manage these objects we rely on Leaflet [2], a JavaScript library designed for the creation of interactive maps, capable of handling multiple layers of information and that stands out for its extreme lightness, an essential requirement in working with GIS since the amount of data is usually several orders of magnitude larger than the application itself. The interaction with the Application Programming Interface (API) is carried out through Leaflet, adding to the map layer, a new layer with the markers that are used to signal the position and other attributes of the educational institutions and the communal boundary polygons.

In order to give context to the application, the geographic information layers displayed in the frontend are represented on geographic maps of Chile, obtained from OpenStreetMap [32], given that Leaflet easily integrates with its Application Programming Interface (API).

Finally, the data obtained through the API is shown with one of the standard libraries in graphic representation for JavaScript, D3.js [7]. This library extends, using JavaScript, the

---

[1]You can find its website at https://s3e2.cl/app and its GitHub repositories for the frontend and backed at https://github.com/itsshompi/s3e2-app-reactjs) and backend (https://github.com/itsshompi/s3e2-backend, respectively.For the latter, it is necessary to request access to the authors.
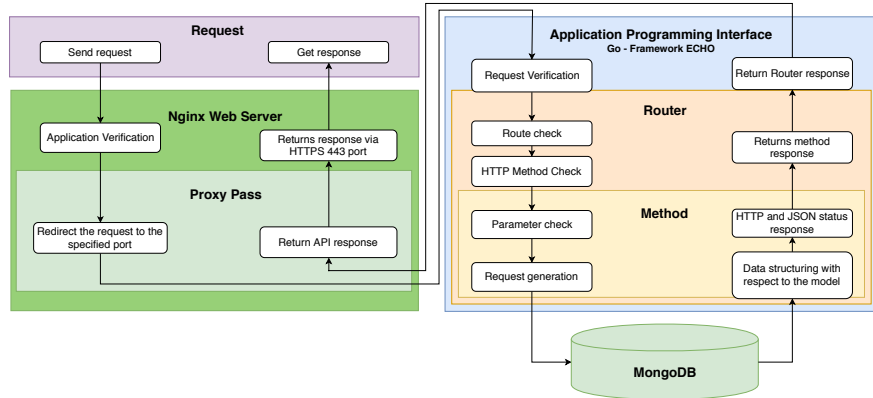
Figure 2: API Architecture.

SVG, CSS, and HTML formats to plot the data that one of its functions receives as input. The graphics produced are SVG files created *on-the-fly*, that is, they are vector diagrams that allow their modification in a scalable and straightforward way to the user's visual system and its interaction with the platform.

2.2   Application Programming Interface (API)

The API is one of our architecture's vital services, due to the need for an expedited connection between the data and the frontend. Therefore, and although it is not indeed a free software tool (but open-sourced and freely licensed), it was decided to implement the API in Go [19], the backends' programming language designed by Google. Go is a compiled, concurrent, and straightforward syntax language, which allows not only for high computing performance but also a shorter implementation time.

As each call outputs a JSON formatted string, the API is, thus, of the REST (*Representational State Transfer*) style; for this reason, we chose to use ECHO [28] as the Go implementation framework. In ECHO, API calls take the form of routes that are parsed using an optimized HTTP router, which, by selecting the appropriate method, queries the database, and returns the appropriate output. ECHO's philosophy means that HTTP errors are handled centrally, which allows adapting not only the outputs but their logging and the functions to be executed when they appear as well. Finally, ECHO allows to automatically scale the application, which gives high performance in environments with JSON type outputs[31].

2.3   Data

The data used in this web-based GIS were obtained from the Open Data of the Study Center of the Ministry of Education of Chile [29], specifically from the databases for researchers section. Of all the available data, we use the following:

- Directory of schools from 2008 to 2018.

- Summary of basic education enrollment in schools from 2008 to 2018.

- Academic and socioeconomic data of students from 2008 to 2018: The socioeconomic data of interest corresponds to the provisions of Law 20.248, on Preferential School Grant

Table 1: Indices created by data collection.

Collection	Unique Indices	Multiple Indices	Total
Students	7	11	18
Institutions	5	5	10
Communes	3	1	4

Table 2: Space occupied by the documents and indexes of the collections and average times to access all their records with and without indices.

Collection	Documents	Indices	Avg. Index	Avg. without Index
Students	8.2 GB	1.9 GB	502.8ms	13047ms
Institutions	64.3 MB	7.9 MB	5ms	137ms
Communes	3.4 MB	65.5 KB	1ms	12ms

(Subvención Escolar Preferencial, SEP): an indicator for preferred students, another for priority students and a field for beneficiary students. This data allows for the classification of the student population into "vulnerable" and "non-vulnerable" classes. From this classification, it is possible to calculate the segregation indices implemented in the platform.

2.4   Document-oriented Database (MongoDB)

Due to the changes that are made every year to the databases of the Chilean Ministry of Education, it was necessary to have a flexible non-tabular data structure. For this reason, the web-based GIS database was implemented in MongoDB, a document-oriented NoSQL database system, i.e., it stores the data in a JSON format [30]. Within the essential features of MongoDB, we can find that it is not necessary to have a previously defined data storage structure, but that these structures form "collections" of documents, where each document becomes a different structure, and fields can be added to or deleted from a document at any time.

Three different collections are defined: students, institutions, and communes. Each of these collections contains different indexes, which provide a significant improvement in the reading speed of the requested data through the different routes of the API. Each collection has been optimized based on the queries that the frontend makes, aiming at improving the latency between the frontend requests and the backend answers.

The indices used for the three collections can be seen in Table 1; in Table 2, it is possible to observe that the creation of these indexes increases the size of the database by approximately 25%. However, when stored in RAM during server load, data access is significantly faster.

2.5   Static Files (Shapefiles)

Shapefiles [15] are data structures that allow storing the information of different attributes and non-topological geometry for the spatial characteristics of a database. An entity's geometry is stored as a set of vector coordinates, while the attributes are stored in a dBASE database.

For this web-based GIS, shapefiles of Chile's communes belong to the collection of the
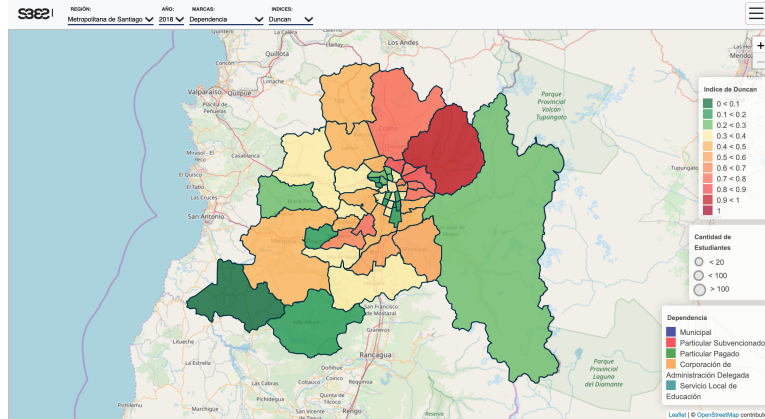
Figure 3: S3E2 graphic display system.

Library of the National Congress of Chile [11], specifically to its territorial information section. These files were stored on a static file server and are served to frontend through an HTTP (GET) query to the API.

3   Results

The S3E2 platform delivers an interface with two main panels, on top and to the left side. The top panel (see Figure 3) allows for the selection of the region of Chile, the year of interest (at this time, from 2008 to 2018), values associated with the schools shown (dependence, total students enrolled, rural/urban condition) and one of the segregation indices of a set of those most used in literature (Duncan [25, 13], Atkinson [3], Isolation [4], Entropy [35], Mutual Information [17], Gini [25]).

As shown in Figure 3, after selecting a region the map will automatically be positioned, using a window with the appropriate zoom level to visualize all its communes, from the corresponding *shapefiles*. The bottom layer presents the map of the displayed territory, obtained from OpenStreetMap. When selecting a year, a query will be made to the API, which will deliver the value of the selected index for each of the communes in the region, producing a color change associated with a predefined scale: greens, for low segregation, oranges, for high segregation, and reds, for excessive segregation.

When clicking on a commune, the view changes to zoom in, displaying its basic schools. Each one is displayed as a circular mark, varying in size according to the enrollment. A translucent color represents the variable selected in the upper panel for the schools (dependency, either municipal, private subsidized, fully private, and a couple of special kind of schools; and rurality condition of the school).

The left panel will also appear, displaying the general data of the selected commune: its name and the value of the selected index. Through buttons, it is possible to display a graph that shows the time series of the commune for total enrollment and for the number of priority students enrolled (see Figure 4).

Selecting one of the commune's schools, its information for the year chosen in the top panel is shown in the lower part of the left panel: name, dependence, rurality condition and totals for enrolled, priority and preferential students (according to conditions of vulnerability stated in
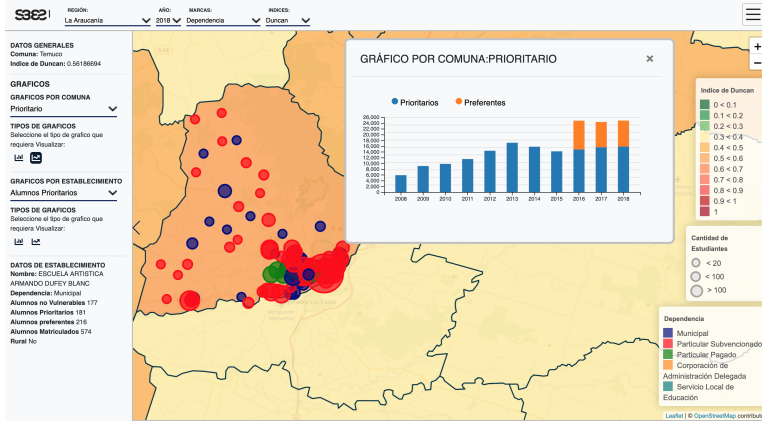
Figure 4: S3E2 graphic display system.

Table 3: Comparison in access to web-based GIS platforms.

Platform	Lighthouse Scoring				
	Perf.	Acc.	BP	SEO	PWA
Chile's Census (ESRI ArcGIS)	36	67	50	82	2/7
Jefferson, TX (Sidwell—Portico and ESRI ArcGIS)	34	84	57	67	5/7
Orange County, CA (Timmons Group)	88	75	64	78	4/7
S3E2	100	91	79	90	5/7

the Law 20.248) and non-vulnerable students (see Figure 4). By pressing buttons that appear on top of this information, the time series of priority and total students corresponding to the selected school are shown.

Using Google Lighthouse v3.0 [20], an open-source, automated tool for in-depth analytics of web pages, it is possible to measure web-based GIS quality. Results are presented in five different dimensions: Performance (Perf), Accessibility (Acc), Best Practices (BP), Search Engine Optimization (SEO) and Progressive Web App (PWA). As shown in Table 3, in a non-comprehensive comparison with some web-based GIS sites, S3E2 has better scoring than the other three applications tested in all dimensions. Concentrating on performance, S3E2 obtained 100 out of 100 possible points; compared to the website of Chile's census (based on ESRI's ArcGIS) or the website of Jefferson County in Texas (partially based on ESRI's ArcGIS) there is a mean difference in performance of 65 points, which shows that our customized solution is substantially faster than the most commonly used one in different environments.

4   Conclusions

This paper has presented S3E2, a web-based GIS for the visualization and analysis of socioeconomic segregation in Chile's primary education system. Although the segregation is calculated at the communal level using widely known indices in specialized literature, it is possible to study their components at different levels: regional, communal, and institutional. This web-based GIS was carried out, almost entirely, with free software tools (except for the Go language which, as mentioned, is open-sourced and freely licensed), which not only lowers its cost but

also allows for the customized control of the development in a careful and detailed way. Hence, it consumes very few resources, both on the client and the server, unlike generic tools for commercial use, which apart from being expensive are usually very demanding in computational resources. Customization has some other advantages that are clear in this case, by using the scoring in the tests implemented with Google Lighthouse v3.0. Given S3E2's architecture, we believe that implementing a web-based GIS within some other context is feasible with minimal modifications.

As future work, geostatistical analysis components will be incorporated to visualize the spatial autocorrelation of segregation. Likewise, we will add mechanisms to calculate the correlation between segregation and the results of standardized tests, to account for some objectives of the CONICYT FONDECYT 11170583 Project of which this web-based GIS is a product. Finally, additional tools for searching and filtering information will be included, as well as for exporting data tables, graphs, and maps for the use of interested stakeholders.

Acknowledgments

References

[1] Daron Acemoglu. Lecture notes for graduate labor economics. Technical Report 14.661, Massachusetts Institute of Technology, Massachusets, 1999.

[2] Vladimir Agafonkin and Leaflet. https://leafletjs.com/. [Last accessed 19-07-2019].

[3] Anthony B. Atkinson. On the measurement of inequality. *Journal of Economic Theory*, 2(3):244 – 263, 1970.

[4] Wendell Bell. A probability model for the measurement of ecological segregation. *Social Forces*, 32(4):357–364, 1954.

[5] Cristian Bellei. El estudio de la segregación socioeconómica y académica de la educación chilena. *Estudios pedagógicos*, 39(1):325–345, 2013.

[6] Michael A. Boozer, Alan B. Krueger, and Shari Wolkon. Race and school quality since brown vs. board of education. Technical report, National bureau of economic research, 1992.

[7] Michael Bostock. D3 data-driven documents. https://d3js.org/. Last accessed 21-07-2019.

[8] Nieves R Brisaboa, Alejandro Cortiñas, Miguel R Luaces, and Matías Pol'la. A reusable software architecture for geographic information systems based on software product line engineering. In *Model and Data Engineering*, pages 320–331. Springer, 2015.

[9] Howard Butler, Martin Daly, Allan Doyle, Sean Gillies, Stefan Hagen, and Tim Schaub. RFC 7946: The GeoJSON format. Technical report, The Internet Engineering Task Force, 2016.

[10] David Card and Jesse Rothstein. Racial segregation and the black–white test score gap. *Journal of Public Economics*, 91(11-12):2158–2184, 2007.

[11] Congreso Nacional de Chile. Biblioteca del Congreso Nacional de Chile. https://www.bcn.cl/. Last accessed 19-07-2019.

[12] Françoise Delannoy. *Education Reforms in Chile, 1980-1998: A Lesson in Pragmatism*, volume 1 of *Education Reform and Management Publication Series*. Banco Mundial, 2000.

[13] Otis Dudley Duncan and Beverly Duncan. A methodological analysis of segregation indexes. *American sociological review*, 20(2):210–217, 1955.

[14] Gregory Elacqua and Rodrigo Fabrega. *El consumidor de la educación: El actor olvidado de la libre elección de escuelas en Chile*. PREAL, Santiago de Chile, 2004.

[15] ESRI. Qué es un shapefile. http://desktop.arcgis.com/es/arcmap/10.3/manage-data/shapefiles/what-is-a-shapefile.htm. [Last accessed 21-07-2019].

[16] Facebook Open Source. https://es.reactjs.org/. Last accessed 19-07-2019.

[17] David M Frankel and Oscar Volij. Measuring school segregation. *Journal of Economic Theory*, 146(1):1–38, 2011.

[18] Kostas Gobakis, Aggeliki Mavrigiannaki, Kostas Kalaitzakis, and Dionysia-Denia Kolokotsa. Design and development of a web based gis platform for zero energy settlements monitoring. *Energy Procedia*, 134:48–60, 2017.

[19] Google. https://golang.org/. Last accessed 19-07-2019.

[20] Google. Lighthouse scoring guide. Last accessed 13-09-2019.

[21] Jonathan Guryan. Desegregation and black dropout rates. *American Economic Review*, 94(4):919–943, 2004.

[22] Eric A Hanushek, John F Kain, and Steven G Rivkin. New evidence about Brown v. Board of Education: The complex effects of school racial composition on achievement. *Journal of Labor Economics*, 27(3):349–383, 2009.

[23] Akhlaque Haque. GIS, public service, and the issue of democratic governance. *Public Administration Review*, 61(3):259–265, 2001.

[24] Caroline Hoxby. Peer effects in the classroom: Learning from gender and race variation. Technical report, National Bureau of Economic Research, 2000.

[25] Julius Jahn, Calvin F Schmid, and Clarence Schrag. The measurement of ecological segregation. *American Sociological Review*, 12(3):293–303, 1947.

[26] Richard Kingston, Steve Carver, Andrew Evans, and Ian Turton. Web-based public participation geographical information systems: an aid to local environmental decision-making. *Computers, environment and urban systems*, 24(2):109–125, 2000.

[27] Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación and Oficina Regional de Educación para América Latina y el Caribe. Informe de Resultados del Tercer Estudio Regional Comparativo y Explicativo (TERCE). *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 14(4):9–32, 2016.

[28] Labstack. https://echo.labstack.com/. Last accessed 19-07-2019.

[29] Ministerio de Educación de Chile. http://datosabiertos.mineduc.cl/. Last accessed 19-07-2019.

[30] MongoDB Inc. https://www.mongodb.com/. Last accessed 19-07-2019.

[31] Fajar Nurdiansyah. Echo vs Iris vs .Net Core benchmark. https://medium.com/geekseat-development-blog/go-vs-iris-benchmark-f7822f84e86f, 2017. Last accessed 21-07-2019.

[32] OpenStreetMap Fundation. https://www.openstreetmap.org/. [Last accessed 19-07-2019].

[33] Ismael Puga. Escuela y estratificación social en Chile: ¿cuál es el rol de la municipalización y la educación particular subvencionada en la reproducción de la desigualdad social? *Estudios pedagógicos*, 37(2):213–232, 2011.

[34] React-Redux. https://react-redux.js.org/. Last accessed 19-07-2019.

[35] Henry Theil. *Statistical decomposition analysis; with applications in the social and administrative sciences*. North-Holland, London, 1972.