

Teaching statistics through the Real Estate Data Analyzer software

Summary

This paper presents the operation of the educational software, Real Estate Data Analyzer, to teachers and future teachers. This software has been conceived for the Spanish educational system, however, due to its characteristics, it can be internationally used. It is a specific software that allows secondary school students have a different vision of the subject, with which they can carry out complete statistical analysis projects on the characteristics of the buildings that surround them.

Keywords: Teaching statistics; Improving classroom teaching; Secondary Education; Educational Software.

Introduction

The use of educational software is the most appropriate resource to help students carry out statistical projects. When working with students using projects, the theoretical knowledge about statistics and its application as well as other areas of knowledge come to light simultaneously; there is a need to apply technical and strategic knowledge to increase students' motivation towards statistics by working with real tasks and data. The use of an educational software allows the organization and analysis of the data and performing the calculations accurately. In addition, it brings support for small investigations in the field, since students can focus all their attention on thinking, reasoning and solving problems (Dolan, 1979; Hall, 2011).

An educational software chosen by a teacher must be objectively evaluated to affirm that it is suitable for the purpose. In this work, we understand the evaluation of educational software as the measurement of the quality of this type of resources through a series of indicators (Marquès, 2002) with the purpose of guiding users and, in particular, teachers in the pedagogical adequate software use.

The main purpose of this work is to present the REDA software features. It is a powerful tool that allows, in addition to analyzing the Spanish real estate market, teaching and learning statistics from continuously updated real data and through a high capacity of graphic representation.

The Real Estate Data Analyzer (REDA) software

The Spanish real estate market is very important in the economy of the country, so its study is of great relevance for different reasons. Due to the high dynamism of this market, it is necessary to have accurate and immediate information.

The software designed and introduced in this work has been created as a resource for the teaching and learning of statistics, with the objective of showing its potential and benefits in a fully contextualized way to students by real practical cases and in a graphic way.

Through its use, students can create projects with the following purposes:

- Extract information on the characteristics, both qualitative and quantitative, of homes, premises, offices and garages from any point of the national geography, which are for sale or for rent at that moment, based on geographical coordinates and a radius of search.

- Save the obtained data to be able to make future updates that show the properties that have been sold or rented, as well as new available offer.
- Sort data according to a specific characteristic or filter the results so that only those that meet a certain condition are displayed.
- Classify each variable into qualitative and quantitative.
- Make graphs for each of the selected variables, such as bar charts, pie charts, histograms or box plot diagrams. Also, the different graphs obtained by dividing the sample according to the different modalities of a qualitative variable can be visualized.
- Calculate descriptive measures of the quantitative variables for the whole or for the different subsets constructed for the different modalities of a qualitative variable.
- Obtain the geographical distribution of the properties analyzed in a satellite map, by including points on it with a color graduation that offers information on the analyzed property.
- Save all the results in HTML format or copy them in a word processor.

Through the obtained results, students will have clear and rigorous information of the real estate market that surrounds them, created by them, through a simple and intuitive program.

Results

Program characteristics

The Real Estate Data Analyzer (REDA) software has been developed in JAVA language to allow its use in different available platforms. Calculations and statistical graphs elaboration are done through a link between the program and the engine of R. In addition, the acquisition of data requires an internet connection.

Therefore, the recommended hardware requirements are: 2 GHz processor or higher, 2 GHz RAM - 4 GHz if the analysis of a large number of data will be carried out - and internet connection.

The software requirements are: operating system compatible with the R software - which are practically all - Java 1.8.0_40 or higher, R 3.2 or higher and RJava library installed. These last requirements are necessary to generate the link between the program and the R engine, which makes the installation process difficult to some extent. This has been solved by writing of a step-by-step installation guide of the program and all the necessary elements included in the installation package.

The application is composed, mainly, by three systems: data acquisition and storage of the information system, user interface and information analysis system; these are interconnected to allow the real state data base use and management.

The data acquisition system is formed, in turn, by three layers: application -which is responsible for the information processing-, database - in charge of the real estate data storage-, and the set of functions given by the company - called API - to obtain real estate information from their server to which we have access under a collaboration agreement.

The user interface allows the visualization of the data collected in the data acquisition process, as well as its filtering and sorting, it can also be exported as CSV files to be used in other calculation or statistical programs.

The information analysis system is responsible for dealing with the information stored through the data acquisition system. This system is based mainly on three layers: the program logic, which communicates with the user interface; the JRI library, and the R engine.

Operation of the program

Prior to the use of REDA, teachers must work with students on the concept of geographic coordinates, as well as obtaining them, for any area, through applications and websites such as Google Maps.

Teachers must also work on the real estate market basic concepts, such as types of real estate that exist, as well as the differences in the use thereof, different existing real estate operations, etc.

When the program starts, students access the main window, shown in Figure 1. As it can be seen, the menu consists of five buttons through which they can: create a new project, open and manage an existing project, update its data in the case that a project has been previously opened, perform filters or orderings and carry out the desired statistical analysis. If students try to access any of the last three buttons without having selected a project, the program will automatically redirect to the corresponding window, indicating that it is necessary to open a project before.

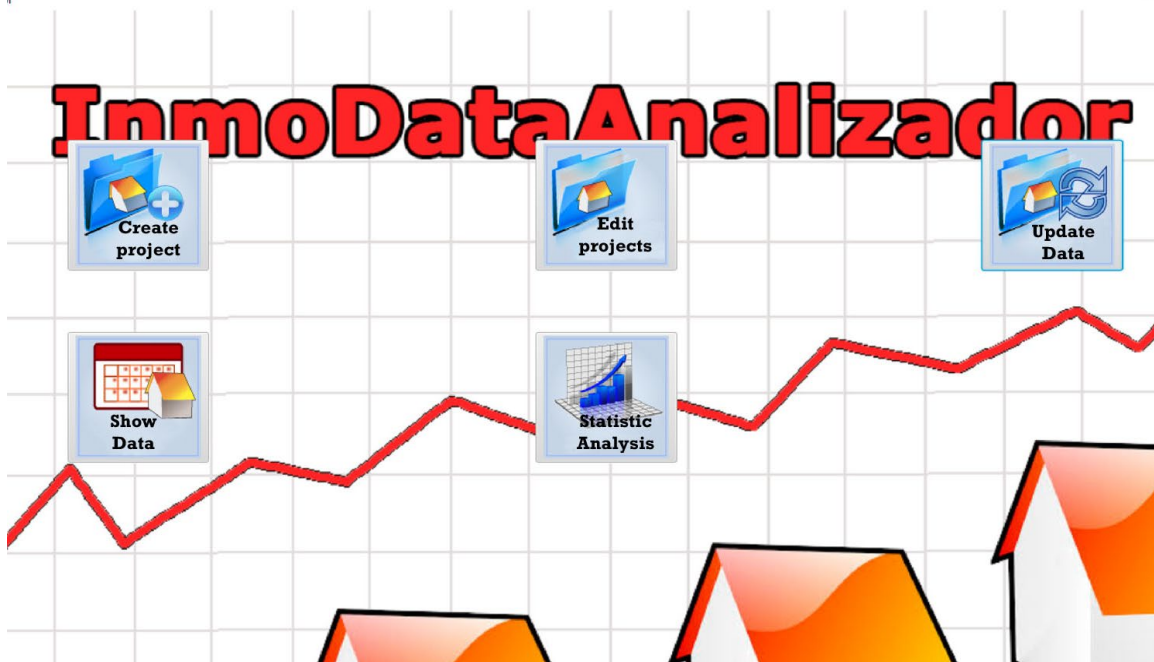


Figure 1. Main window of the program

First, students must access the project creation window, which starts with the search of the properties on offer in a certain area from any point of the national geography. The program will search for information related to buildings contained in a circle from which the center is defined, by means of its geographic coordinates, and the radius, expressed in meters. (Figure 2)

Figure 2. Project creation window

In addition to these parameters, they must select the type of property they want to analyze, choosing between housing, commercial premises, offices and garages; and the type of offer, for sale or rent.

Students can also perform filters on the search, in order to obtain only properties with a specific minimum or maximum price or size.

Once the search has been carried out, the program will show the main characteristics of the selected properties. The most relevant are those indicated in Table 1.

Table 1. *General characteristics of the real state*

Real state location	Real state characteristics	Ad characteristics
Length	Type	Release date
Latitude	N. ° Rooms	N. ° photos
Distance	N. ° Toilets	URL photos
Neighborhood	Size	URL
District	Floor	
Municipality	Price	
Province	Price mm/dd/yy	

There are specific variables of each property that can be requested, and that will show relevant information about the property. These variables are automatically enabled for selection when students set the property for which they want to carry out the study.

These variables are, for each type of property, the following:

- Housing: garage, elevator, swimming pool, storage room, drying rack, air conditioning, fitted wardrobes, and terrace; with the information of whether or not each of them is available. The State field is also available with three possible values: good condition, reformed or new building.
- Commercial premises: Air conditioning and smoke outlet; dichotomous variables with the information of whether or not they have them, the corner field, indicating whether the location is at the intersection of two streets, and the location field that gives information on whether the location is at street level, whether it is part of a shopping center or it is located in a mezzanine.
- Office: Garage, Independent hot water, Elevator, Air conditioning, Independent heating and Security; indicating whether the property has these characteristics or not. The Distribution field gives the information on whether the office is open-plan or if, on the contrary, it has a distribution divided into rooms or partitioned. Finally, we have the Use field that refers to the building in which the office is located. This can be exclusively dedicated to offices or, on the contrary, it can also contain dwellings.
- Garage: Security and automatic door; with information about its availability. It also has the motorcycle parking field, for the case that the garage is intended for this type of vehicle because of its size.

Once the search fields are completed, it starts with just pressing the Recover data button. When doing so, and after making a web query, a preview is shown indicating the number of properties found, some characteristics of the first 50.

At this time, students must decide between saving the project or return to the previous window to make changes, depending on the obtained results. To save the project, it is as simple as to indicate the name the student wants to assign to it. This will be stored along with the rest of previously created projects, which can be accessed from the Manage Projects button in the main window.

The previously generated projects are listed in this window, where we can select the one we want to work with or just delete those projects that we do not want to keep stored.

It is important to note that the project only contains the specifications of the search at this moment. To obtain the properties contained in it, simply go to the project update window and click on the Recover data button.

From this same window, each project can be updated later so that new additions or deletions in the list or variations in the prices of existing properties can be studied, which will be indicated in the Price / mm / yy variable.

Once the information has been obtained, it can be consulted by clicking on the Show data button in the main window. From this window, students will be able to review the complete result of the search for the first time.

By clicking with the left mouse button on the title variable, students will be able to organize the data from that variable upwards. A new click performs sorting downwards. The order is alphabetic if the variable is of qualitative type.

If, on the other hand, the right button is clicked, a menu is displayed and shows a list of modalities of variables that can be enabled or disabled to only display the desired type of properties.

A button in the lower right corner allows to export the information to a CSV format file, so that it can be accessed by other programs such as spreadsheets. The created file will only contain the real estate selected by the student after using the filter.

One of the fields shown in the listing is the ad URL, which includes the functionality that when the address is clicked, the browser defined by default opens and goes to the real estate ad on the website of the Idealista.com portal, where photographs and description of the property can be consulted.

After visualizing the obtained data, students can perform the statistical analysis through the Statistical study button on the main window. Clicking on it, the program accesses a window that is composed of different tabs in which many statistical studies of different variables can be made: univariate, bivariate and regression analysis, and satellite maps representations. To access each of the tabs, just click on the desired option in the left menu.

The first window allows a previous data filtering, in the same way as in the viewing window, which will be applied to the statistical study that is carried out.

The second one allows a complete univariate descriptive study. First, the variables are divided into two groups: qualitative and quantitative.

Different descriptive measures such as mean, median, minimum, maximum, standard deviation, variance or coefficient of variation can be calculated for the second one. There can also be calculated confidence intervals for the average, for a given level of confidence, for students of more advanced courses.

The available graphic representations are the bar diagram, the sector diagram, histogram and box diagram. For all of them, you can choose between representing them from absolute and relative frequencies.

If, in addition to a quantitative variable, students select a qualitative one, descriptive studies will be carried out for each of the categories of the qualitative variable chosen. An example of the result obtained for an extensive study of the offer of homes for sale in the city of Cordoba, in which average and standard deviation of the price of real estate have been calculated according to their type, as well as the different histograms of relative frequencies, can be seen in Table 2 and **Figure 3**

Once the desired study has been configured, by pressing the Configure button, the program internally sends the necessary information to the R engine, which gives the results of the study so that they can be displayed by the program through HTML format.

Table 2. Frequency table of the price of real estate according to their type

	Chalet	Cottage	Duplex	Flat	Penthouse	Study
Size	1226	40	59	4213	131	9
Mean	259,7953	181,9500	125,2881	99,2644	123,5115	94,5556
Typical deviation	225,5575	171,1696	49,6922	40,7628	54,9676	69,4858

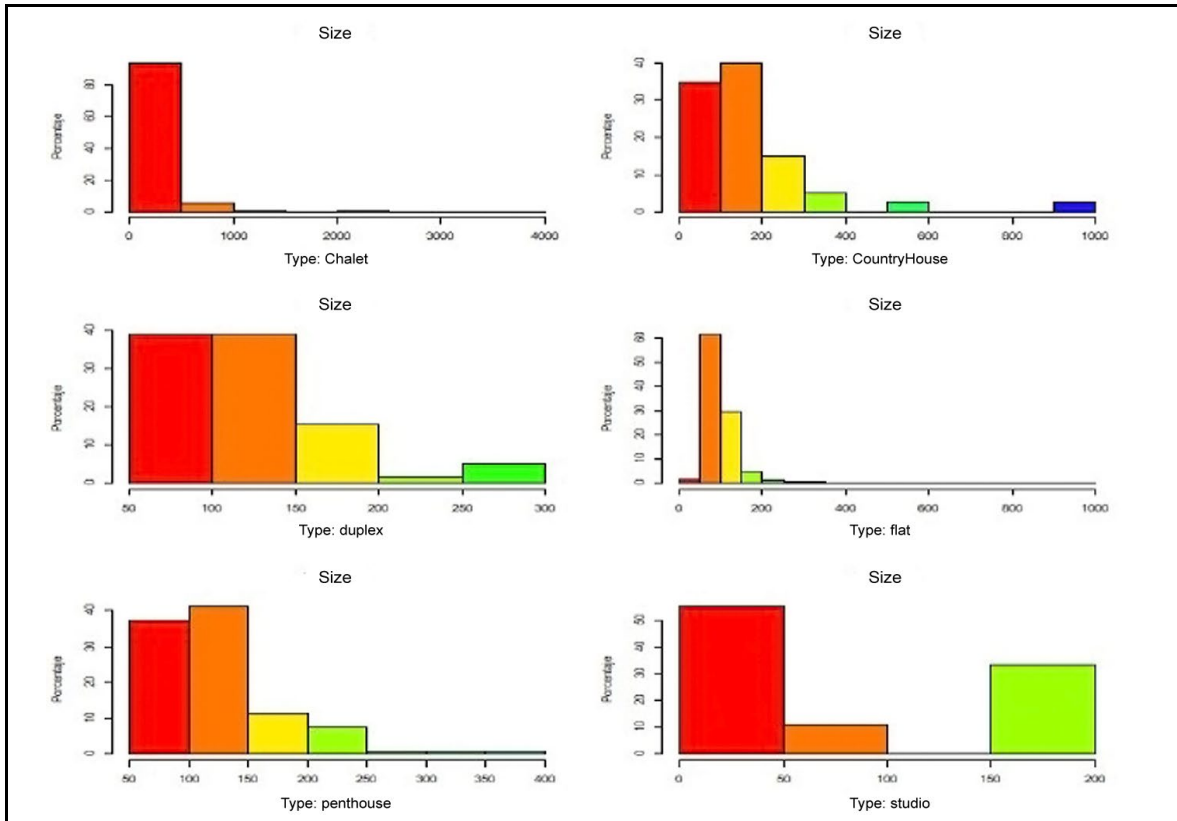


Figure 3. Histograms of exploratory study

Taking the number of variables and options available in this tab into account, there is a large number of combinations of studies that can be carried out and be of great interest.

To contextualize the students' experience even more, they can prepare a satellite map in which, besides locating the analyzed properties with points, different color graduations show the value of a variable as well as operations between some of them.

This can be done through the Map tab, shown in Figure 4. For its elaboration, the requested map of the Google Maps repository is automatically downloaded, it is sent to R, where the points are added from the selected configuration.

In this tab, the variables are divided again between qualitative and quantitative, since the map elaboration differs between one and the other. In addition, different types of maps can be selected such as physical map and satellite street map; the zoom of the map or the size and opacity of the shown points can be modified too.

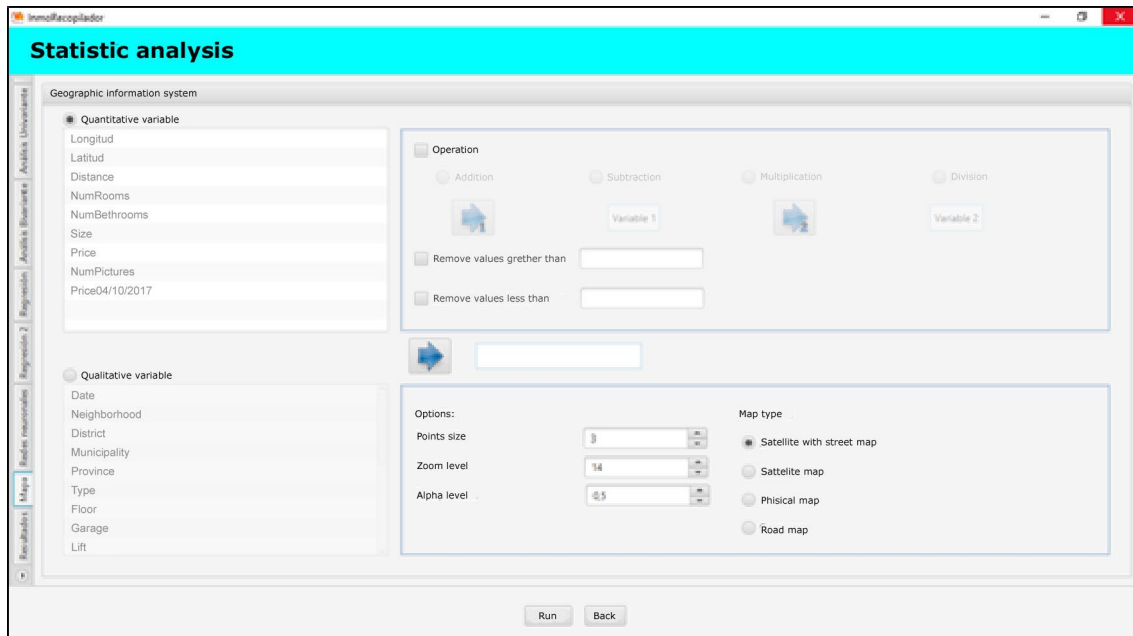


Figure 4. Map tab

For the quantitative variables, the points that locate the properties are colored through a gradient between yellow and red, so that the smallest values of the variable are represented by yellows, and the highest values with red. In addition, maps can be built for a variable value or for a simple operation between two of them. That is, you can build a map according to the price, size, or price per square meter - price / size.

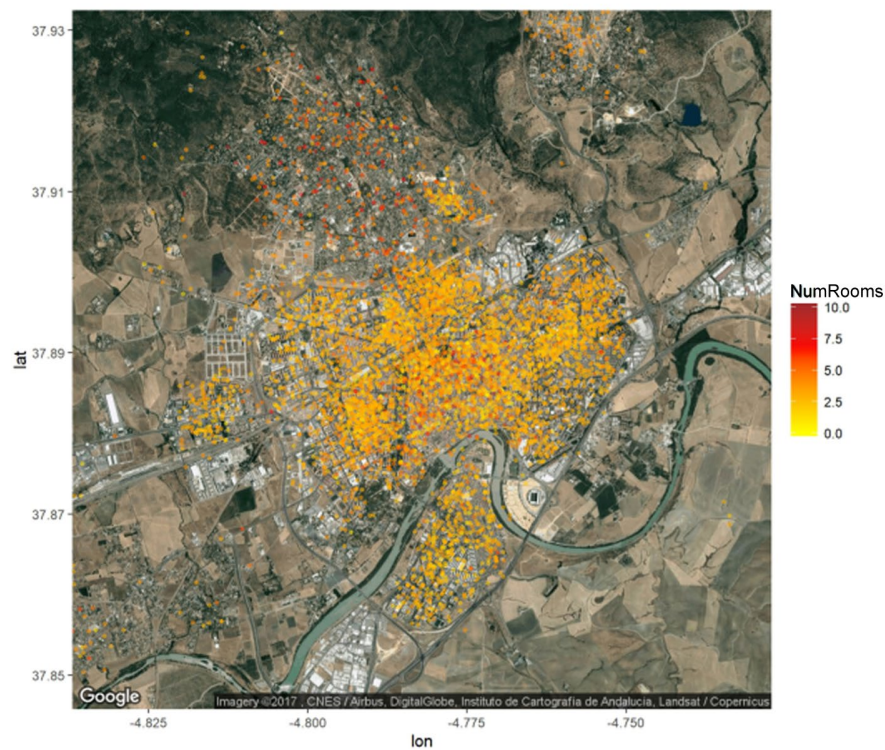


Figure 5. Satellite map of the city of Cordoba.

Figure 5 shows the map resulting from placing the properties for sale in the city of Cordoba at a specific time, as well as the information regarding the number of rooms.

For qualitative variables, points of as many colors as categories are represented by the variable. These colors are provided randomly for each requested study.

Program evaluation

As an evaluation tool for the REDA program, the simplified catalog and evaluation of educational programs of Marqués (2002) was used, as mentioned above.

An explanatory session of the software was carried out, which included the software installation process, a detailed explanation of the operation of the program, and the subsequent realization of exercises to apply it to real cases.

Due to the small study sample and the qualitative nature of the answers, the analysis of the results obtained will be only descriptive, indicating the main evaluations given to the different items.

Regarding the pedagogical and functional aspects, didactic effectiveness, relevance of learning and contents, didactic versatility, application/creative approach and promotion of self-learning stand out positively. The less positive evaluations were given to the ease of installation and use, because the link between JAVA and R requires a laborious installation process, and access problems consideration.

In relation to the technical and aesthetic aspects, high or excellent valuations were obtained in most of the items, so it can be said that the software is original and makes use of advanced technology, it is reliable in its execution, it allows interaction with the activities and its contents are well structured. However, the audiovisual environment obtained a somewhat lower rating.

Regarding the cognitive effort required, among the answers given unanimously are understanding / interpretation, analysis / synthesis, search / assess information, reasoning and exploration / experimentation.

Finally, in terms of global assessment, highlights that 5 rated the software as excellent, while 4 gave it a HIGH rating.

Discussion and Conclusions

The software introduced in this paper was designed as a tool for the analysis of the real estate market in Spain. The main objective is to bring Secondary Education students closer to the possibility of carrying out this analysis on real data of the environment that surrounds them, in a way that motivates them in its use and moves them away from the monotonous and sequential mode with which in most of the occasions is taught statistics.

This tool allows students to make real searches of the properties that are for sale or for rent at any point of the national geography, in particular those available in their closest environment, place it on a satellite map and perform a thorough analysis of their characteristics.

The multidisciplinary panel of experts selected to evaluate this tool considers it capable of motivating students by influencing the usefulness that the subject has over the surrounding environment.

As negative aspects, it is worth noting the difficulty in installing the program, something that is currently being worked on, which is motivated by the complexity of the different technologies it uses and its connection. Another aspect that should be deepened lies in the graphic interface of the program, which should improve two aspects, on the one hand, give a more pleasant and appropriate appearance according to the age of the student, and secondly, provide new features for students with special curricular needs.

Finally, as a future line of action, we intend to evaluate the influence of the use of REDA software in the teaching of statistics of secondary school students.

References

- Dolan, O. (1979). Learning Statistic Through Project Work. *Teaching Statistics*, 1(2), 34-41.
- Hall J. (2011) Engaging Teachers and Students with Real Data: Benefits and Challenges. In: Batanero C., Burrill G., Reading C. (eds) *Teaching Statistics in School Mathematics- Challenges for Teaching and Teacher Education*. Dordrecht: Springer.
- Marquès, P. (2002). Evaluación y selección de software educativo. *Revista Comunicación y Pedagogía*, 185, 31-37.
- Mills, J. D., & Raju, D. (2011). Teaching statistics online: A decade's review of the literature about what works. *Journal of Statistics Education*, 19(2), 1-28.