Analysis and Visualization of Big Data (2022), **1**, 1–4

OPEN DATA JOURNAL

MINI ARTICLE

How Catalunya's Future Evolves: A Geographical Analysis of 6th grade *Competències Bàsiques*

Arnau Codina,* Júlia Vicens, Aleix Nicolás, Ignacio Prieto, and Irene Valladares

Universitat de Barcelona
*Corresponding author. Email: arnaucodvil@gmail.com

Abstract

This project investigates the performance of students in Catalonia across the standardized tests known as "competències bàsiques" in Catalan, Spanish, English, and mathematics. The primary focus of the study is to examine whether students' geographical background influences their test results and whether this result is also dependant of temporal aspects. Our findings reveal notable disparities in student performance, attributed to both geographical and temporal factors. Particularly, students in rural areas exhibit lower test scores compared to their urban counterparts. We come to the conclusion that these differences may stem from inadequate funding in the education sector, a challenge exacerbated in rural zones. This study sheds light on the urgent need for a change in education funding in order to target and eliminate these inequalities.

Keywords: Competències Bàsiques, Catalunya, Education, 6th grade, Geographical

1. Background

The data that has been used throughout this project recollects the grades of Catalan students of the 6th primary grade in the standardized *competències bàsiques* exams and can be found in the open Catalan site *dades obertes de Catalunya*[2]. The data set provides students grades in these standard tests in different subjects from 2009 to 2022. Moreover, it also provides the geographical zone each student lives in so it is possible to analyze this data by comparing the grades in each student for each different geographical area. In this project we want to find out whether the geographical background of the students will affect in any way the results they will have in different subjects. One of the more interesting aspects of the project will be studying how students from different zones score in exams of Catalan and Spanish as it is thought that, usually, people from rural areas have a better knowledge of Catalan but struggle with Spanish. We also want to see whether grades vary significantly from one year to

another and whether this would be due to variance in each generation's capabilities or due to an increase or decrease in the exam difficulty.

2. Methods

The analysis of the data was done in python. We used the pandas, seaborn and matplotlib libraries and the code can be found on our GitHub repository [1]. The code consists on two main parts. The first part consists on downloading the wanted data and format it in order to facilitate its study. The first step is separating our data by geographical zone and year as these are the parameters we want to study. In order to be able to study its evolution we compute the mean over all students for a given year and zone for each of the subjects we will study. With this we can obtain a pandas data set that includes the geographical zone, year and averaged grade for all subjects. By pandas melting this data set we obtain the wanted format in order to proceed.

The second part of the code consists of the graphical analysis of the data. To analyze it we have used the seaborn scatterplot feature that allows us to create a graph for each subject in which we can compare the students' results both throughout the last 12 years and throughout different zones of Catalonia. These plots use a color scale to compare the academic results of each group of students in function of the year, the subject and the geographical zone.

3. Results

Figure 1 shows the results we obtained throughout the study. It shows four colormaps, each of which corresponds with a different subject we were interested in studying, with time and geographical zone as our variables. It is therefore possible to study both the evolution over time of students' performance upon reaching the sixth grade and we can see whether the geographical zone in which they live determines whether they might get better or worse grades.

We can see that most average grades are somewhere between 74 and 76 for all three languages and are way higher for maths. Even if at first glance it may seem all grades are random and not quite dependant on time or zone we can find some interesting outliers in our graphs. It can be seen how, in Barcelona, students tend to score higher than students in other zones and have the highest grades in all four subjects over the period of time we have studied. If we look into it even more, we can see that their performance is especially high in Spanish and English. When comparing for all subjects we see grades in Barcelona outperform the average result by over 2.5 points in Spanish and by 3.5 points in English whereas they only outperform by 1.5 points in maths and by 1.75 in Catalan. This result is somewhat surprising as the exam is the same for every student so no zone should present such consistent overperformances. However, is can be explained as the education system often favours students that live and study in big urban cities rather than those who live in more rural environments as cities offer a better access to diverse studying material. The difference between English and Spanish versus Catalan can also be comprehended if we take into account that, due to migration, many families in Barcelona com from

non-Catalan speaking environments so they perform worse on the exams. This aspect is non as prevalent in less urban places so this differentiation between Catalan and Spanish/English does not appear.

Apart from these geographical differences we can also observe some temporal differences. As the exams should always be of the same level no relevant discrepancies should be seen from year to year for each subject. However, we can see how some years present overall higher grades than others. For example, we can see how in 2012 Catalan the grades are, in average, around 80 to 82 points whereas the year after they plummet back down to around 73. Even if this could be due to a generation having lower capabilities, the homogeneity around the country in scoring higher or lower hints that it most likely isn't a consequence of certain generations being more capable than others. One way of understanding this is that the actual difficulty of each exam varies from year to year.

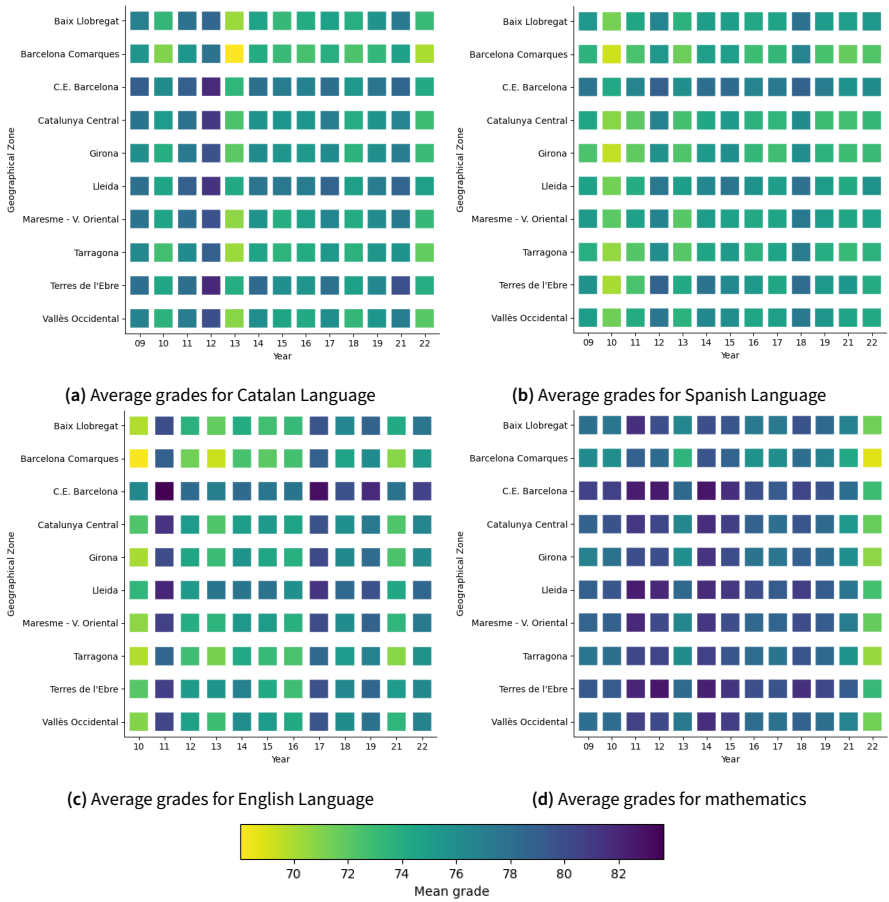

Figure 1. These colormaps allow us to compare averaged grades over all students for the *competències bàsiques* exams for each year and geographical zone

4. Conclusions / Discussion

The main goal of the project was to study whether geographical and temporal circumstances affect academic results in relevant ways. Throughout the study we have seen how both these variable can affect students' performance in academic exams. This can pose several questions on whether these exams are a fair way to compare students from different years as the difficulty of these can also vary from year to year, making most comparisons in average grade futile. Apart from this, this study also suggests that educational inequalities may be present in Catalonia as students from its capital and main urban zone usually scores much higher than those from more rural environments. This could be due to a lack of high quality infrastructures around the country, especially in non-urban locations that receive a lower funding than its capital.

This academic disparity between Barcelona and the rest of the country sheds light on the need of better funding and distribution of educational resources in Catalonia. It is apparent that some zones may have a more challenging time in accessing facilities and infrastructures than may academically help students.

Notes

References

[1] AAVV_JaraySedal. Github repository, 2023. https://github.com/Ronoh97/AAVV_JaraySedal.

[2] d'Educació, D. Avaluació de sisè d'educació primària, 2023. https://analisi.transparenciacatalunya.cat/Educaci-/Avaluaci-de-sis-d-educaci-prim-ria/rk5x-gny6.