2 July - Virtual | 8-11 July - Salzburg, Austria | #useR2024

Transforming Data into Information: Overcoming Challenges in Educational Data Analysis

Natalia da Silva

Motivation

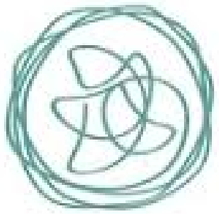
- Learning Management Systems (LMS) or learning platforms like Moodle and Blackboard have become key tools in education. A huge volume of student and teacher data is generated by LMS on a daily basis
- Transforming this data into relevant information for decision-making is a major challenge due to the complexity of the data structure and the difficulty of summarizing the learning process based on it.

Motivation, Plan Ceibal

- One Laptop per Child (OLPC) program, in primary education (2007)
- Extended to secondary schools
- Key role during the COVID-19 pandemic
- Webpage: https://ceibal.edu.uy

National grant

PROYECTO ANII



FundaciónCeibal

Team UDELAR

- Ignacio Alvarez
- Mauro Loprete
- Oscar Montañés
- Jimena Padín
- Bruno Tancredi

FCEA FACULTAD DE CIENCIAS ECONÓMICAS Y DE ADMINISTRACIÓN

IESTA INSTITUTO DE ESTADÍSTICA

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

Lines of Work

1. Tools for evaluating and monitoring the use of Learning Management Systems (LMS) in primary schools in Uruguay. LMS Monitor implemented in a Shiny app.”

2. Identify the key drivers in the use of LMS, understand the sources of variability, and define measures for student engagement.

3. Predict student performance in English tests based on their use of the LMS.

LMS Monitor

Data from LMS

- **DATA**: Students and teachers use of Learning Managment Systems (LMS) from primary schools in Uruguay from 2018 to 2021 (for 4th to 6th grades).

Year	N	Students	Techerss
2018	1.912801	98.054	3801
2019	2.652856	103.168	4474
2020	7.403004	109.019	5477
2021	8.607445	119.065	5299

To transform these data into relevant information for analysis and decision-making it is required to pre-process the data, define relevant summaries, and statistical visualization.

LMS Monitor, ShinyApp

- Should be scalable
- Shuld be modularized
- Should be flexible enough to maintain, modify and test
- Should be reproducible
- Should be prepare to be use for many users at the same time

The app was develop based on `rhino` conected with PostgreSQL data base.

Predictive Models

Academic performance

Goal: Predict English test performance based on LMS student usage.

Different problems:

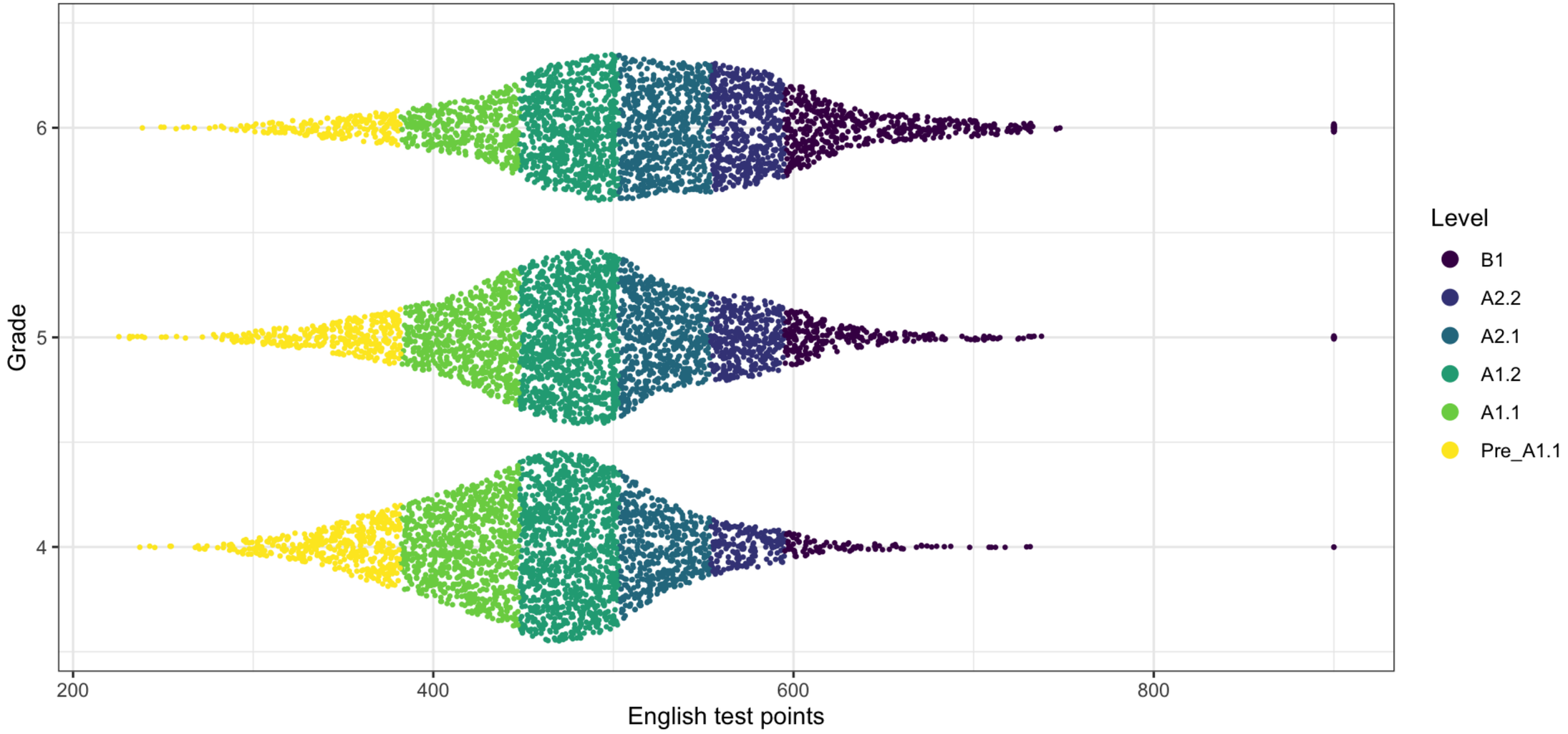
1. Predict the final points in the English test for each student.
2. Predict the average points in the English test at the class level.
3. Predict whether a 6th grade student will reach the A2.1 level as it is expected.

Workflow based on `tidymodels`

Academic performance

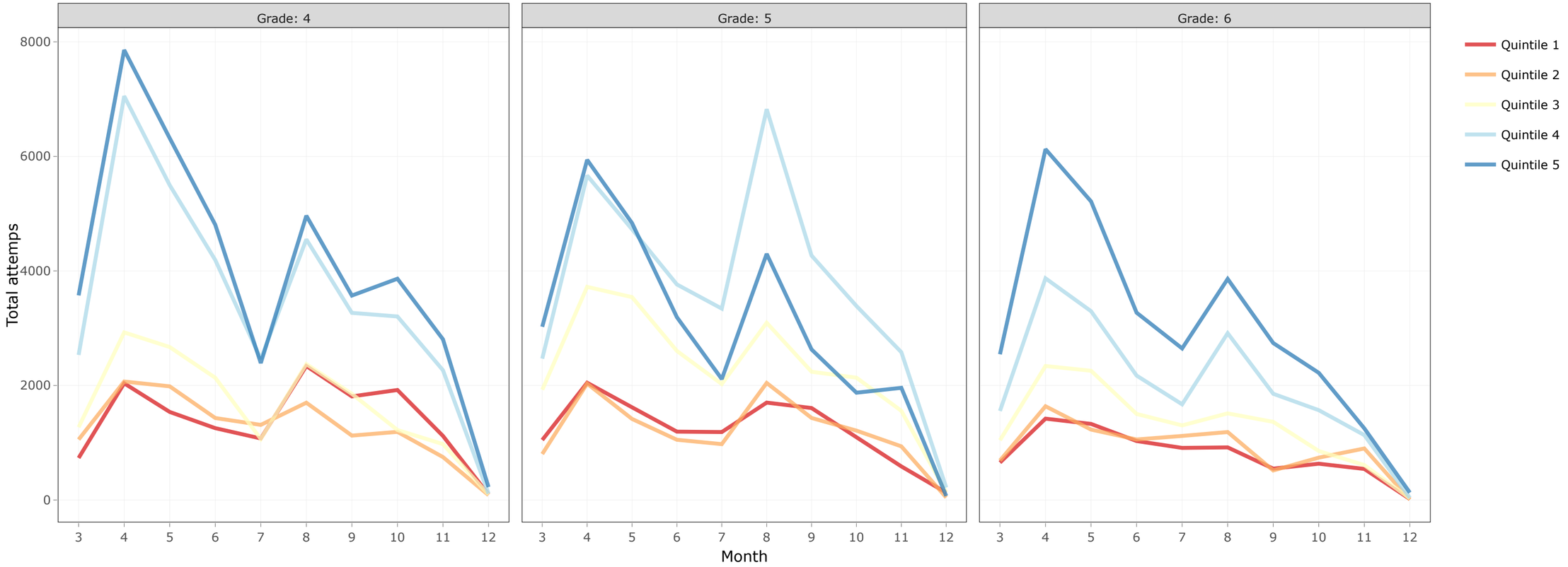
1. Data sources:
    - Performance data: Results from Adaptive English tests
    - LMS data: Usage data from Little Bridge.
2. Focus:
    - Target group: 6th grade students (11 years old).
    - Response: Predict whether students reach the expected English proficiency level (classification problem).

Performance data



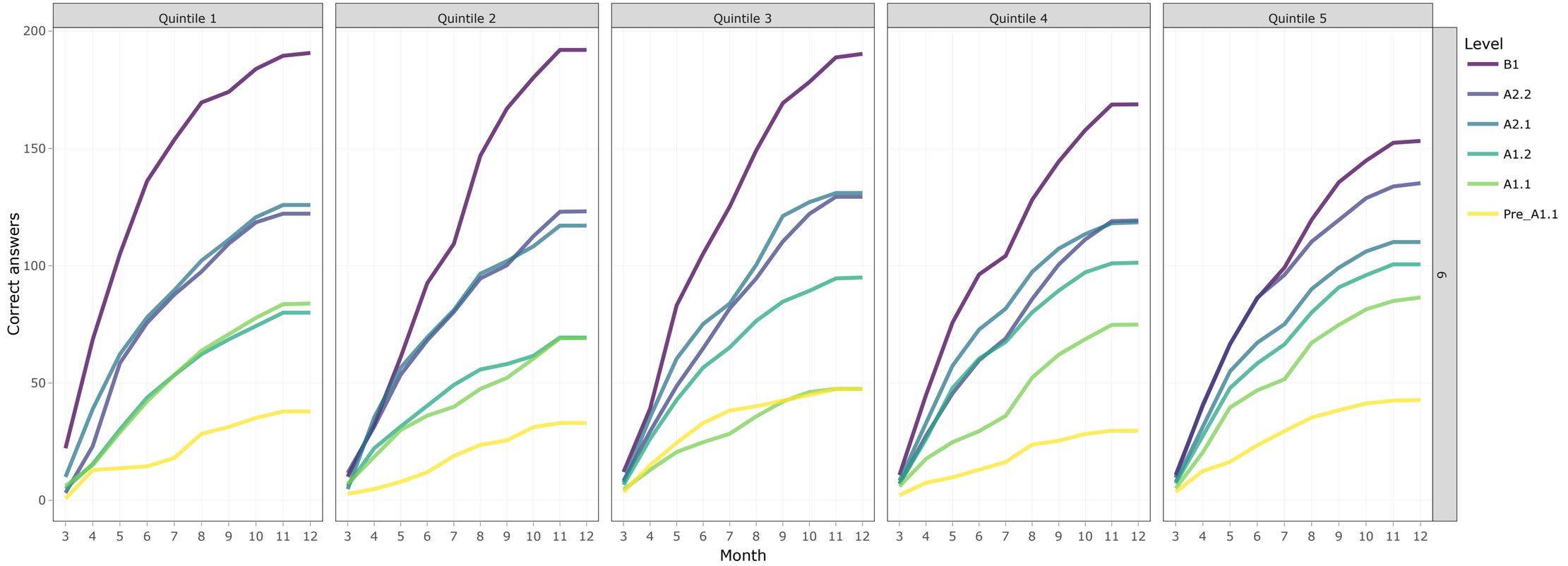
12% of students below A1.1 level

From socieconomic context to LB work



Monthly attemps by Grade and Socioeconomic context

From LB work to academic performance



Monthly right answers by Socioeconomic context and English level

Clasification problem

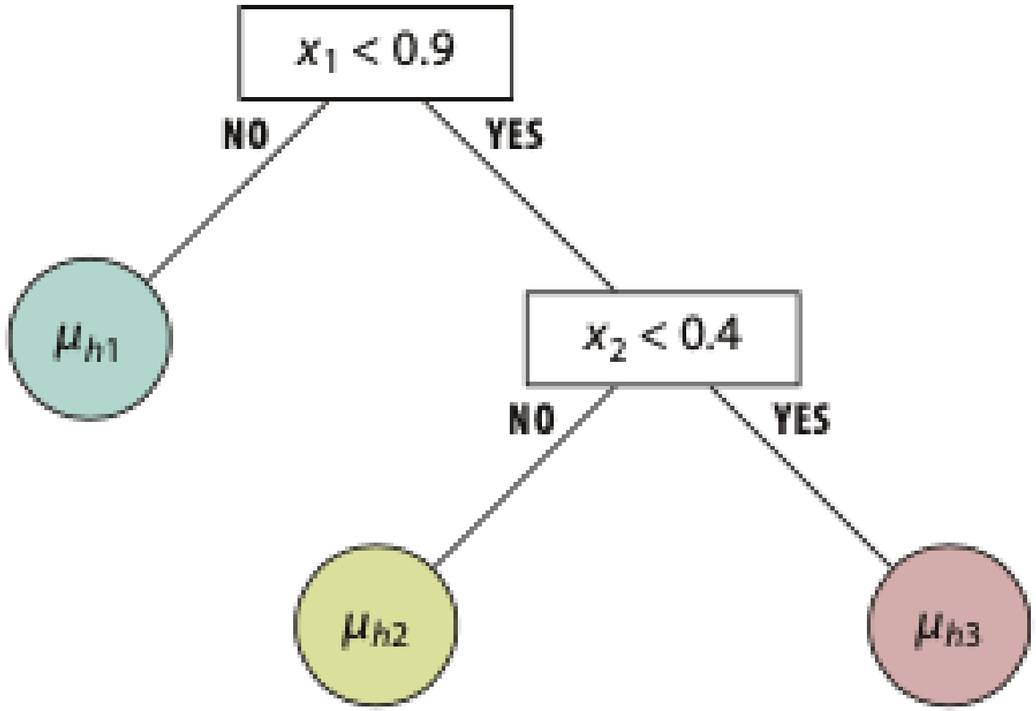
Children in 6th grade are to reach **A2.1** level.

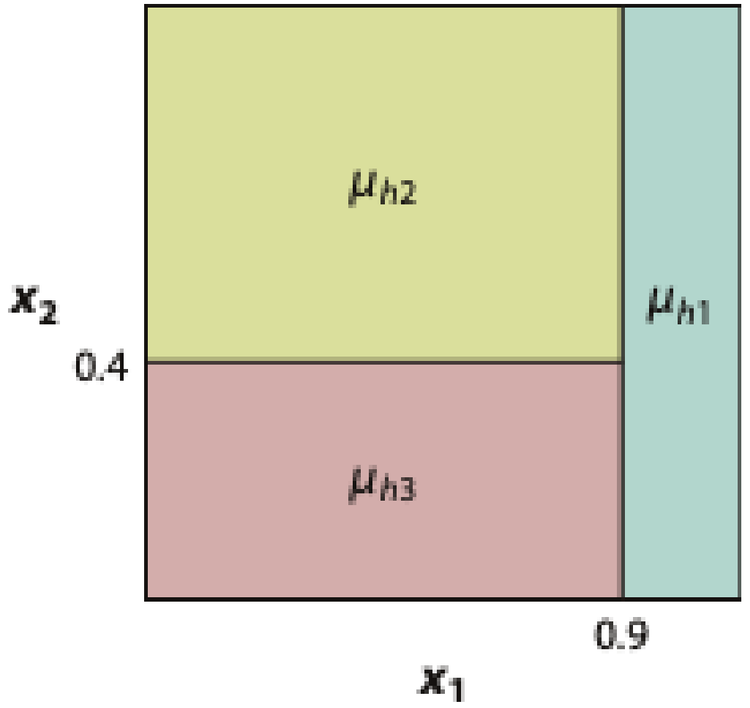
- Response:

$$Y_i = \begin{cases} 1 & \text{reaches A2 level or higher} \\ 0 & \text{otherwise} \end{cases}$$

- Use LB acumulated work up to July
- Fit several statistical learning methods
- Include school random effect, BART

Bayesian additive regression trees





- Picture from: Hill, J., Linero, A., & Murray, J. (2020). . Annual Review of Statistics and Its Application, 7, 251-278.

BART model

A single tree is denoted as:

$$g(x|T, M) = \sum_k \mu_k I(x \in R_k)$$

having two basic parameters: tree structure T and set of leaves values $M = (\mu_1, \ldots, \mu b)$.

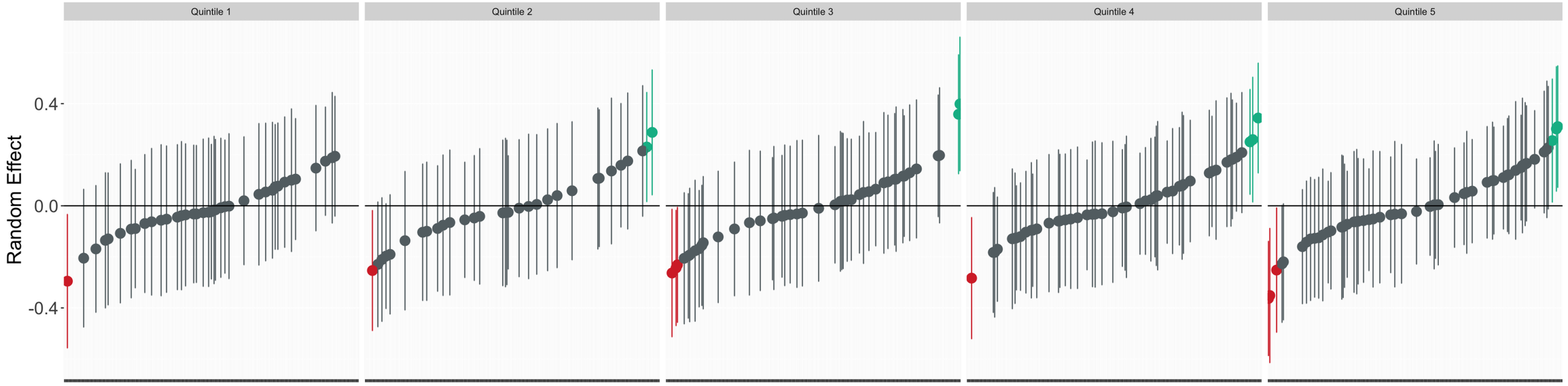
BART: sums of trees model

$$\begin{aligned} Y_i &= f(X_i) + \epsilon_i \\ &= \sum_j g_j(X_i|T_j, M_j) + \epsilon_i \\ \epsilon_i &\sim N(0, \sigma^2) \end{aligned}$$

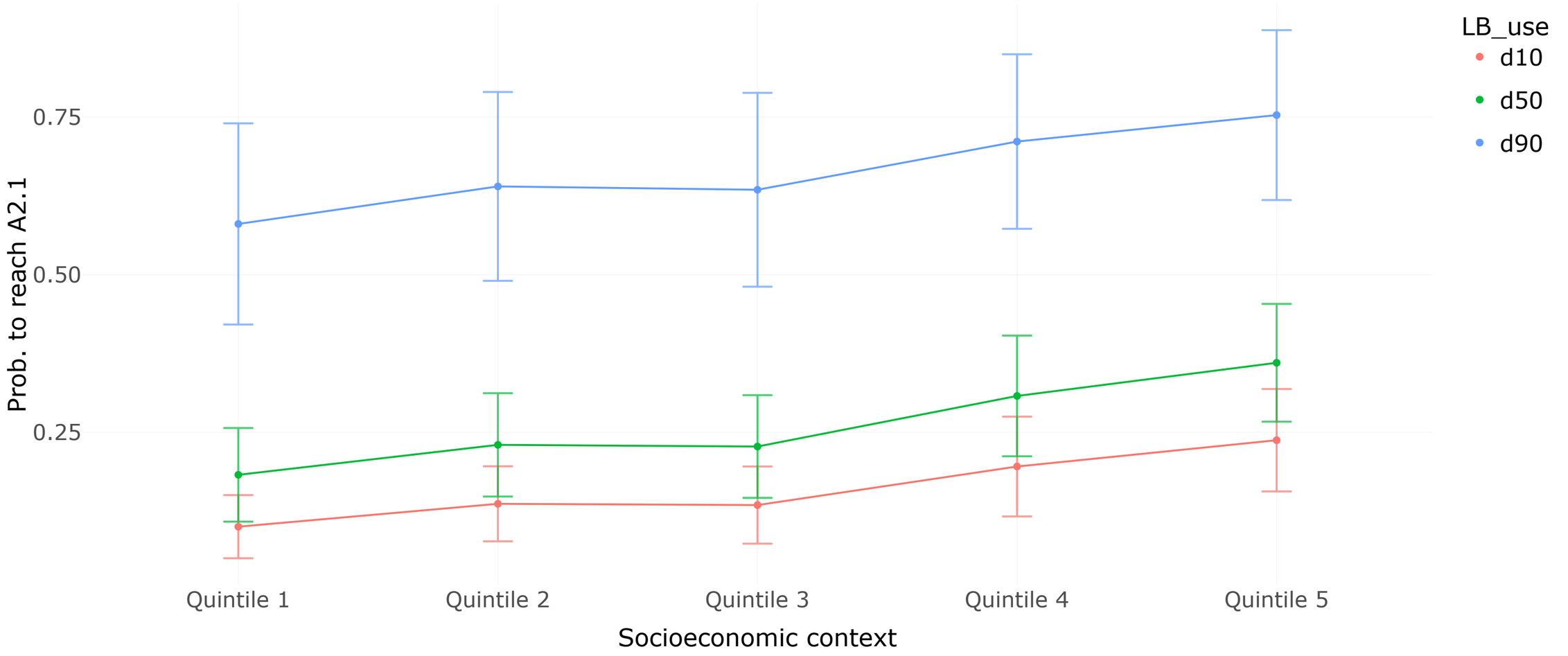
It is possible to add random effects, $f(X_{ig}) + \alpha_g$.

Random effects results

There are schools with positive effetcs in most quintile groups



Predictive model results for some syntetic students



Final Comments

- LMS data can be used to improve education.
- Transforming data into relevant information requires appropriate computational tools and statistical methods.
- We present summary statistical information in a Shiny app based on rhino to monitor and evaluate LMS usage at different levels.
- We built a BART model to predict academic performance at the end of the year.
- Model results can be used as an early warning tool to identify students at risk and centers in need of intervention, as well as centers to learn from.
- Some of the relevant tools used in this project for data wrangling, visualization, and predictive modeling were `data.table`, `shiny`, `rhino`, `plotly`, `ggplot2`, `tidymodels`, `dbart`, and databases (PostgreSQL).

Thank you!