Data Science and Scientific computing - University of Trieste
Data Visualization Final Project

Student performance: a visual analysis

Andrea Barbieri - Francesco Ortu
Anna Pederzani Samarati

A.Y. 2022/2023

Outline

Topic and Dataset

- Our focus: schools and students
- Our goal: visualize links between student's grades, socio-demographic features and external support
- Dataset[1] used: collection of grades and features of 382 high school students from Portugal, gathered through questionnaires and school reports

[1] https://archive.ics.uci.edu/ml/datasets/student+performance

Questions

1. How are the grades distributed among different sexes and study hours?
2. Are "good" and "bad" students different people?
3. How does external support impact the distribution of the Math grades along the academic year?

Math vs. Portuguese: gender-based differences in student performance

This plot displays the correlation between average daily study hours and gender with Math and Portuguese grades. Female students outperform male students in Portuguese, while male students perform better in Math.

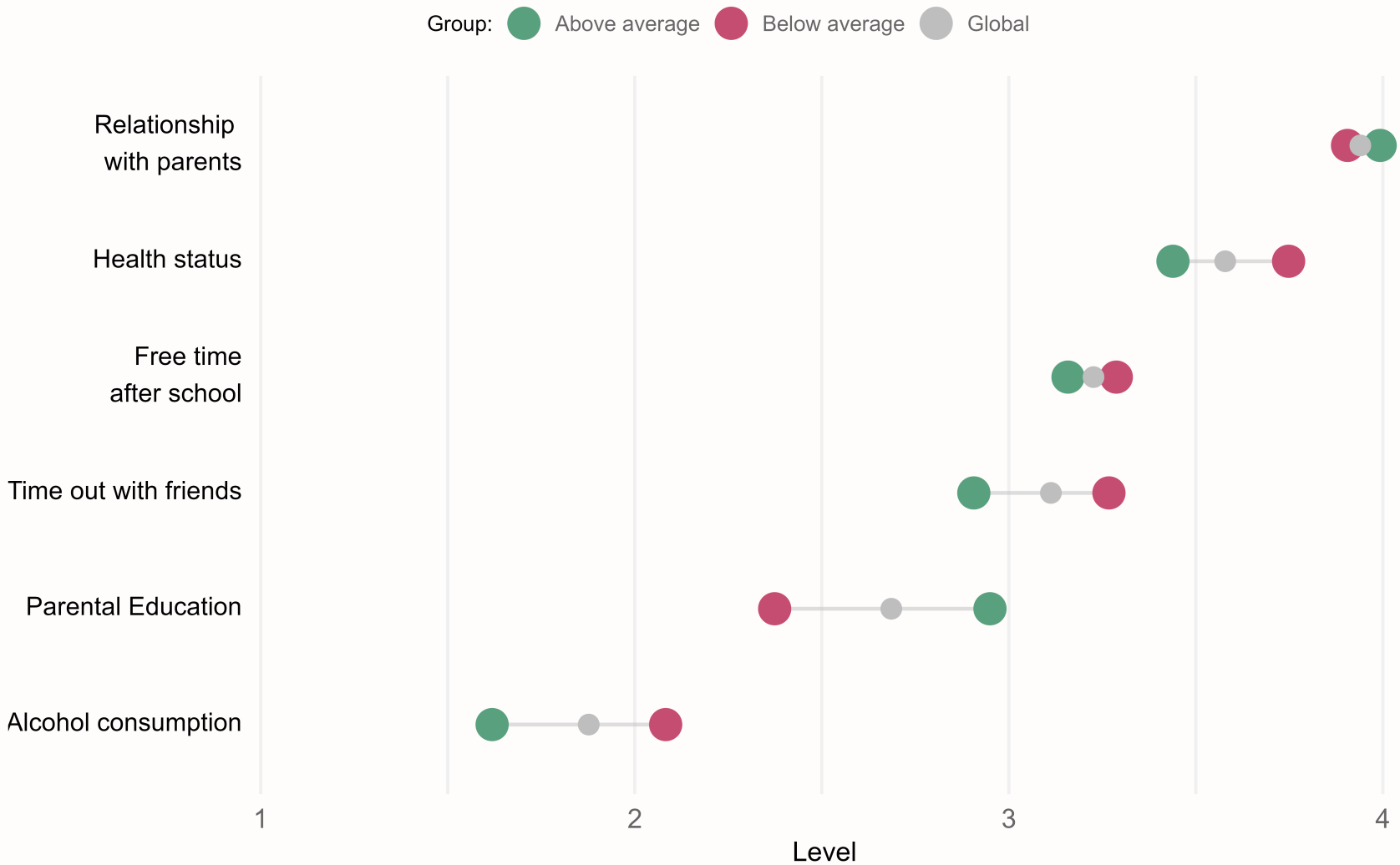
Source: UCI Machine Learning Repository - Student Performance Data Set

Plot 1 - Visual Channels

- Position of circle points: map each student to their grade
- Color hue: gender (red/pink for women and blue for men)
- Saturation: average study hours
- Dashed lines: mean grades

Good and bad students are different people

The following graph compares two groups of students: the ones with a G3 grade in both Portuguese and Math above the global average, 139 students, and the ones with G3 grades below it, 127 students. Each row represents a question whose response ranges from 1 to 5, where the higher the number the better/greater the feature. For each question the average response of the group is reported, together with the global mean.



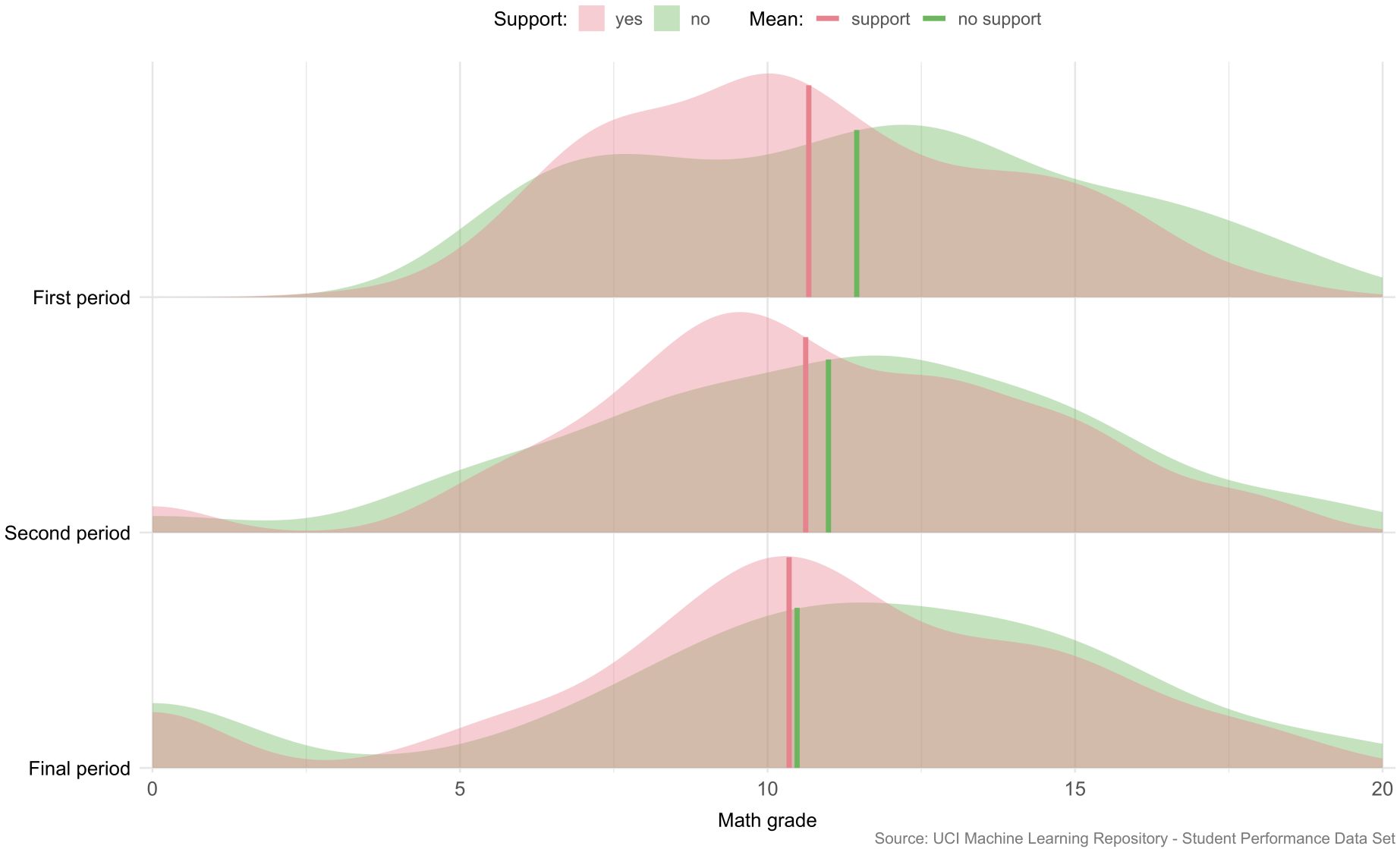
Source: UCI Machine Learning Repository - Student Performance Data Set

Plot 2 - Visual Channels

- Position on a X-axis: average response level of the group
- Position on Y-axis: question addressed
- Color hue: helps distinguish the two groups and the global data
- Points size: highlights relevant data (two groups)
- Length of the lines connecting dots: difference in magnitude between the response of the two groups

Averages of Math grades decrease along the academic year, but support has a positive impact on the drop

The following graph compares Math grades of students with support and without support in three different academic year periods. The means of Math grades of students who do not have support are higher compared to the ones of students with support in all the three periods. The average grades in both groups decline as the school year advances. However, the means of the grades of students with support remain more stable.



Source: UCI Machine Learning Repository - Student Performance Data Set

Plot 3 - Visual Channels

- Vertical levels: three grade periods, ordered by timeline
- Density plots: distributions of the Math grades
- X-axis: Math grade scale (all the three levels are aligned on the same horizontal scale)
- Color hue: to distinguish the two groups of students
- Vertical segments: to map the means of the distributions on the x-axis

Acknowledgments

- **All**: dataset selection, question formulation, plot validation.
- **Andrea Barbieri**: plot for question 2
- **Francesco Ortu**: plot for question 1
- **Anna Pederzani Samarati**: plot for question 3

Thank you for the attention!