Assignment IV

Data Visualization

Benjámin Olivér Maday

2023-11-04

0.1 Task 1

In this assignment, we will be using the nlschools dataset from the MASS package. The dataset is from a a paper titled Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modelling written by Tom A B Snijders and Roel Bosker from the University of Groningen [1]. It contains 2287 observations about eighth-graders in 132 classes in 131 schools in the Netherlands. There are six variables within the dataframe:

Table 1: Description of variables in the nlschools dataset about pupils in the Netherlands

Variable description	Data type
A given pupil's language test score	integer
A given pupil's verbal IQ	continous
Their classe's ID	integer
Class size	integer
Their families socio-economic status	integer
Whether they were taught in a multi-graded class	categorical

Now let's take a look at the relationship between class size, the type of class (whether it's multi-grade or single) and the score the class achieved on the language test.

I used faceting to differentiate between the two types of classes. As we can see on the left side of Figure 1, there is no obvious relationship between the class size and the language test scores, we can only observe a slight decrease as the size gets bigger. In multi-grade classes however the larger the class size, the better the scores get. My hypothesis for this phenomenon would be that multi-grade classes are characteristic in more rural areas, thus a greater class size suggests a larger school and presumably a higher-quality education. Overall, single grade classes still produce better results.

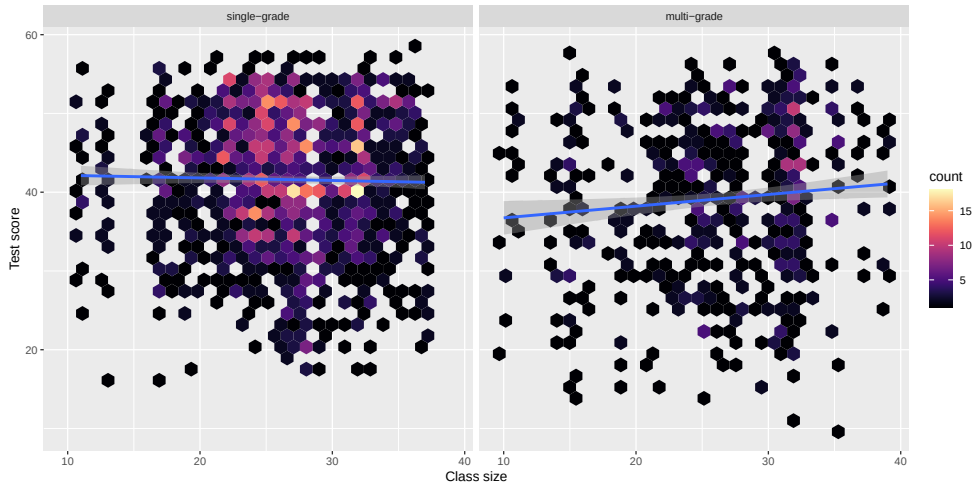


Figure 1: Hexplot of the relationship between the test scores and the class sizes, faceted by the type of class

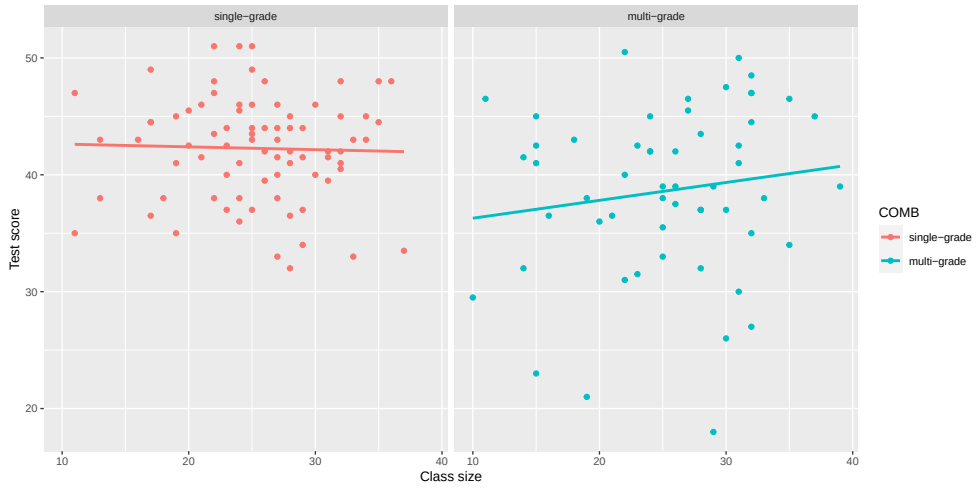
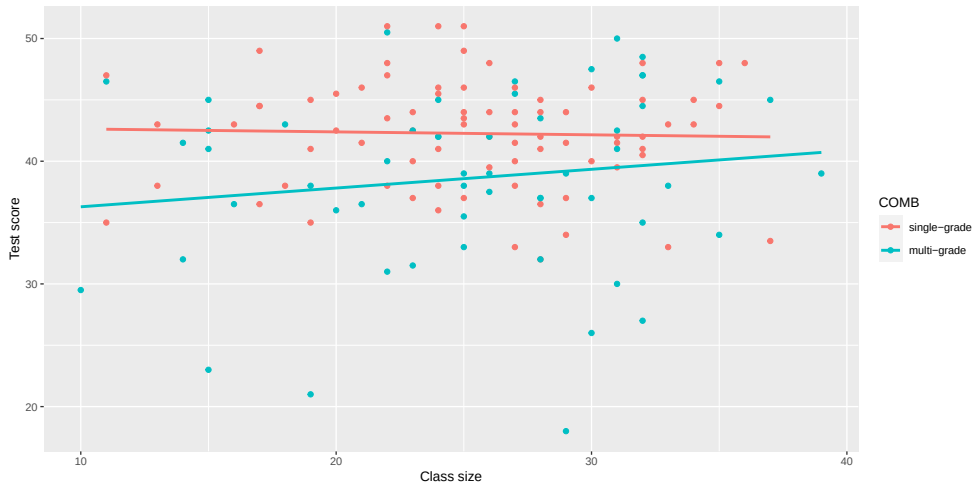


Figure 2: Scatterplot of the median score of students in each class in relation to their class size, faceted by the type of class

0.2 Task 2

The findings from Figure 1 are clearly visible on Figure 2, and the difference between the two class types are more obvious.Another advantage of this summary is that we could now show the two kinds of class types achivements without faceting.



However, it must be said that with this method we lose a lot of individual data points which could be a drawback with other datasets. To provide a brief example: just like in our previous Swiss dataset (where we visualized the religious identity of Swiss regions), there could be great differences in test scores between individuals in a class. Let's imagine a class of 31 students, where 15 students achieved 20 while 16 students attained 50 points on the language test. Taking only the median observation from this class would result in a very deceptive conclusion.

0.3 Task 3

Firstly I wanted to create a categorical variable from the socio-economic status of the pupil's families. In order to do that I took a look at the distribution of them in the dataset, as well as their relationship with language scores.

According to Figure 3 they are quite evenly distributed, thus I created three groups out of them.

We can see that regardless of class type or size, children with a better background attain better academic scores than their less fortunate peers. The class size has almost no effect on single grade classes with students coming from higher income families, but have negative effects in the other two income categories. The class size has a positive effect in multi-grade classes regardless of income levels, but we do not have a lot of observations from this group.

[1] https://www.researchgate.net/profile/Tom-Snijders-2/publication/44827177_Multilevel_Analysis_An_Introduction_to_Basic_and_Advanced_Multilevel_Modeling/links/0c96051ffabd4ca210000000/Multilevel-Analysis-An-Introduction-to-Basic-and-Advanced-Multilevel-Modeling.pdf

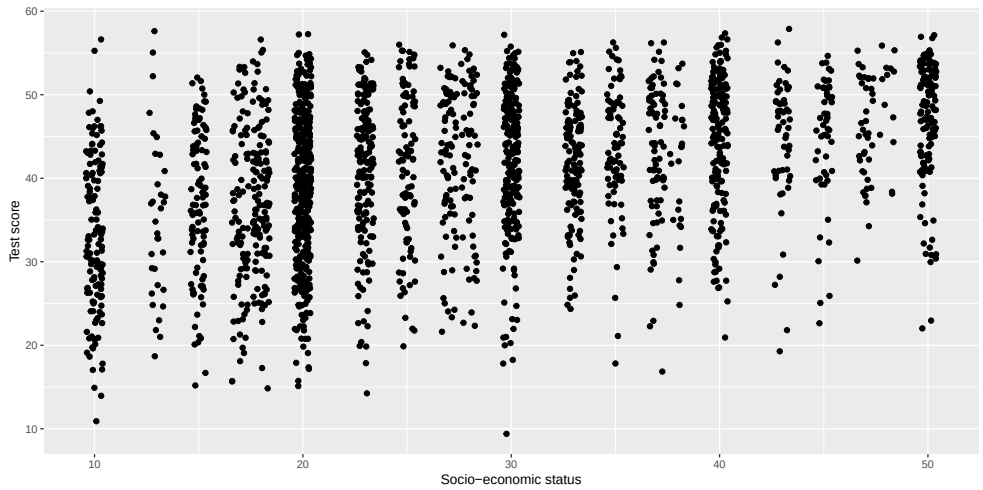


Figure 3: The relationship between a student's family's socio-economic status and their language test score

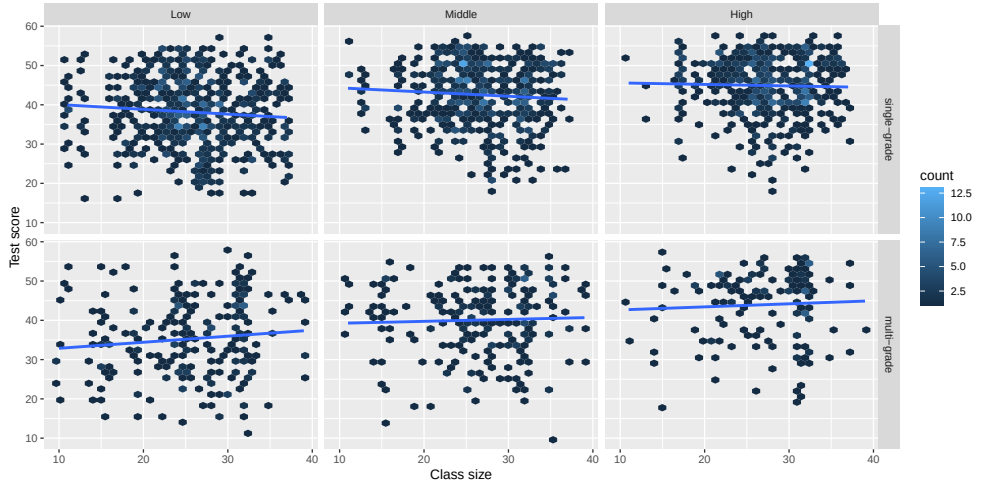


Figure 4: The relationship between a students language exam points and the class' size in which they study, faceted by their familie's economic status and their class type