Diego Valdes

IST 707

Home Work 2

4/9/2019

**Introduction**

America has been wrestling with improving its schools for generations. With Supreme Court decisions like Brown vs. Board of Education that desegregated schools, to the No Child Left Behind act that provided extra funds for the disparity created by financial inequality in schools, the United States has consistently attempted to improve the quality of its educational system. While not every attempt has gone over well or smoothly, as with the implementation of Common Core, there has always been, and will continue to be a consistent effort to make improvements.

There are many factors influencing the quality of schools in America. At the forefront is economic inequality. It is not uncommon to visit two schools minutes apart and derive that the quality of education in the classroom is vastly different. Some schools will have classrooms with multiple computers, while other schools may be lucky to have a few running machines in their media center. This disparity can be found with the quality of text books, school events, and even with the quality of teachers. Anyone bearing witness to such a display of unfairness in America would want to do everything in their power to balance the scales. After all, everyone, regardless of status, deserves the same opportunities.

In the idea of fairness, it is not uncommon to find that entire school districts have attempted to implement a standard curriculum for students. The idea is that it doesn’t matter where a student goes to school, where they live, or who their teacher is, they will experience the same curriculum as the students in the school down the road. Much more than just a list of required topics, districts will write an entire lesson plan, almost a script for the teacher to follow, complete with questions to ask the students and a list of ‘acceptable’ responses. No longer does a teacher have to design their own lesson plan, they can simply follow the instructions. And all the students will get the same education. No matter where they go to school.

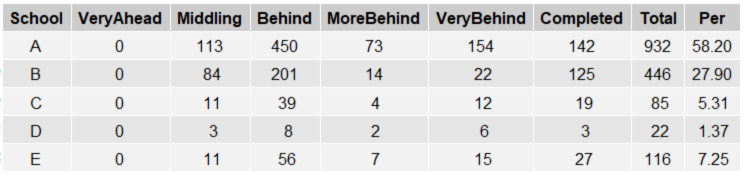
**Analysis**

**About the Data**

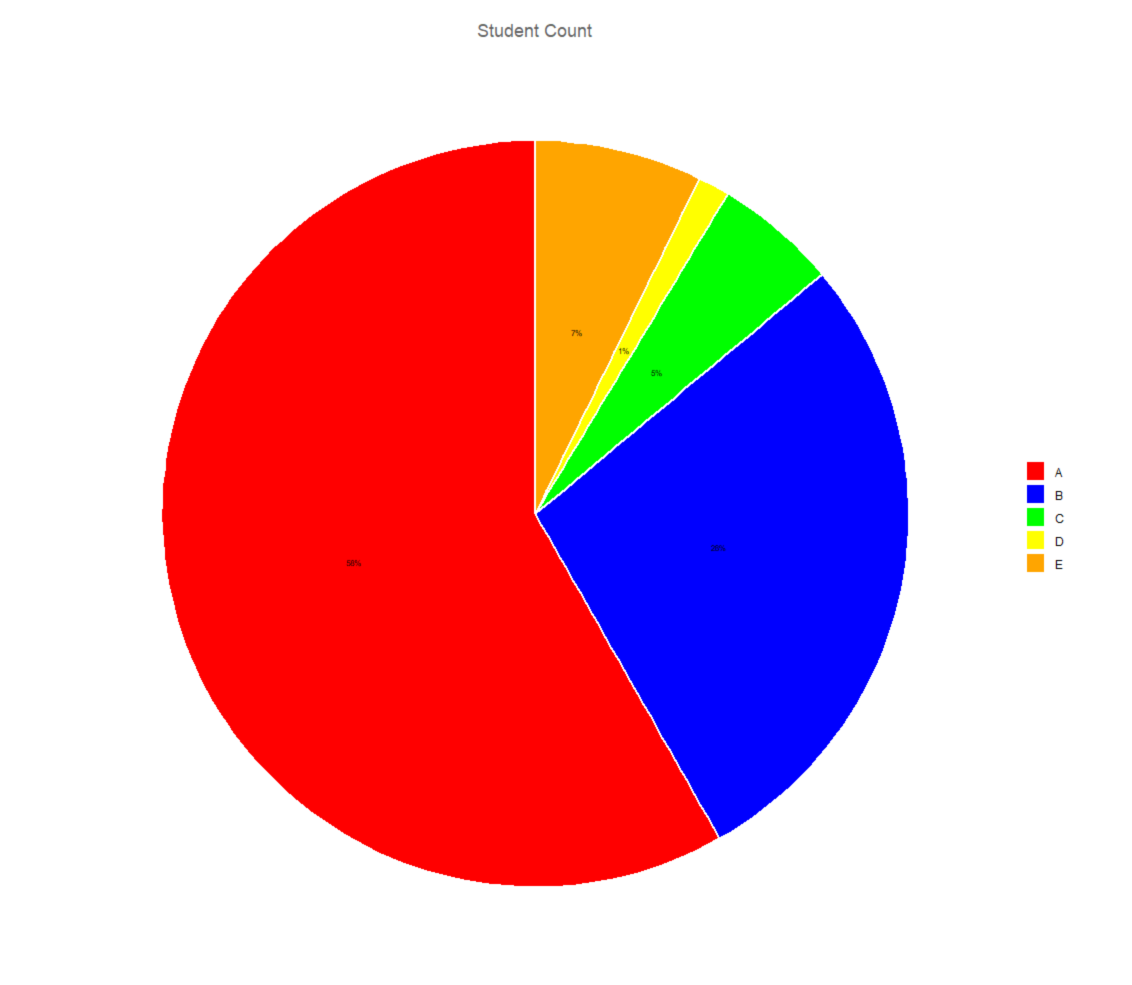
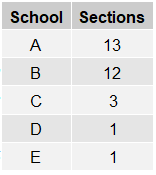
The dataset is the status of students in a new math course being given at 5 different schools. Each school has a certain number of sections (30 sections total), and the students are divided into 6 categories depicting their status through the course, which is 3/4 completed.

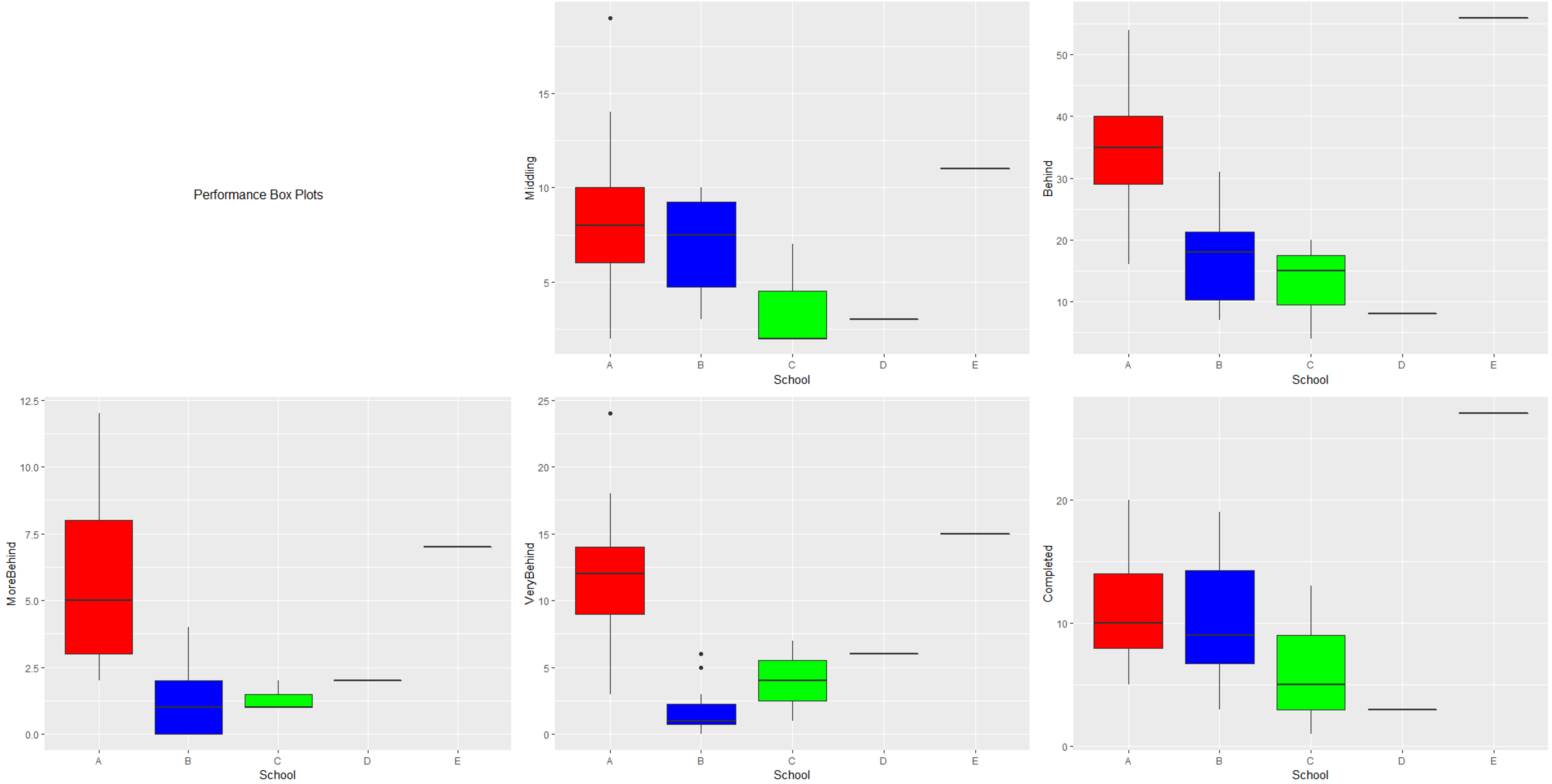
* Very Ahead – more than 5 lessons ahead
* Middling – 5 to 0 lessons ahead
* Behind – 1 to 5 lessons behind
* More Behind – 6 – 10 lessons behind
* Very Behind – more than 10 lessons behind
* Completed – Finished the course

A summary of the data showed that there were no NA or NaN values and there was little to be done to prepare it for analysis. The values for school and section were turned into nominal variables. The data was altered numerous times for the purpose of visualizing and analyzing the data. This included aggregating it by school and section, creating a dataset where the student progress was represented by the percentage of the students, looking at just whether students were progressing well or behind, creating a dataset with scaled values, and creating a dataset to create specific visualizations where the col names (Very Ahead, Middling, …) where turned into values for a column called ‘Status’. The below is an example of aggregating the data to examine the progress by school.



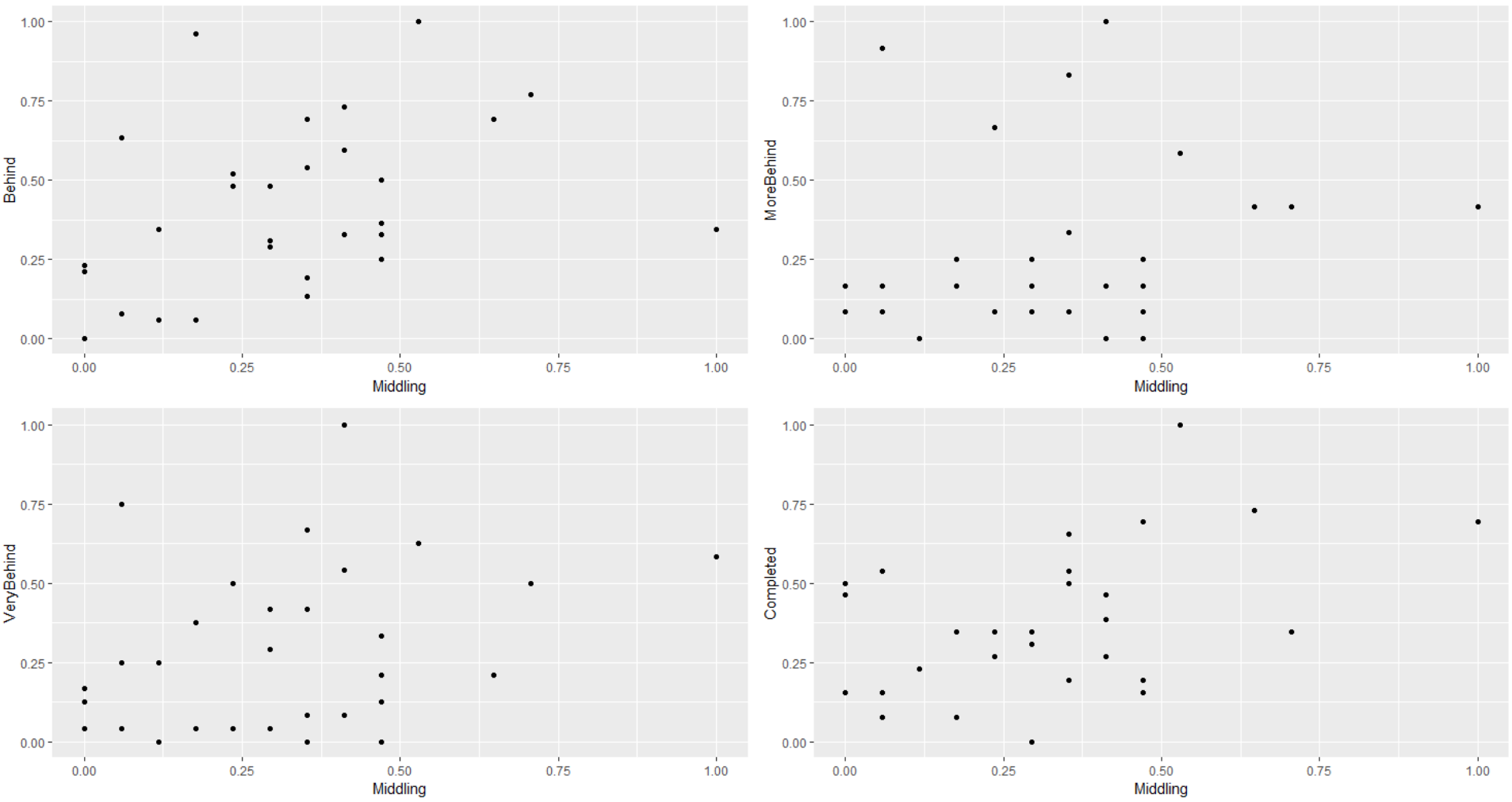
It is important to note that no student fell into the category ‘Very Ahead’ for any school. It is unclear if this is accurate or due to an error. There is also a major difference between the population of each school, the number of sections per school, and the number of students in each section. This can be most seen by comparing School A, 13 sections, 932 students with School E, 1 section, 116 students. These can be highlighted by the below box plots (scaled data) and pie chart (counts), where due to the lack of sections in School D and E, there isn’t enough data to make a fair comparison to Schools A, B, and C.



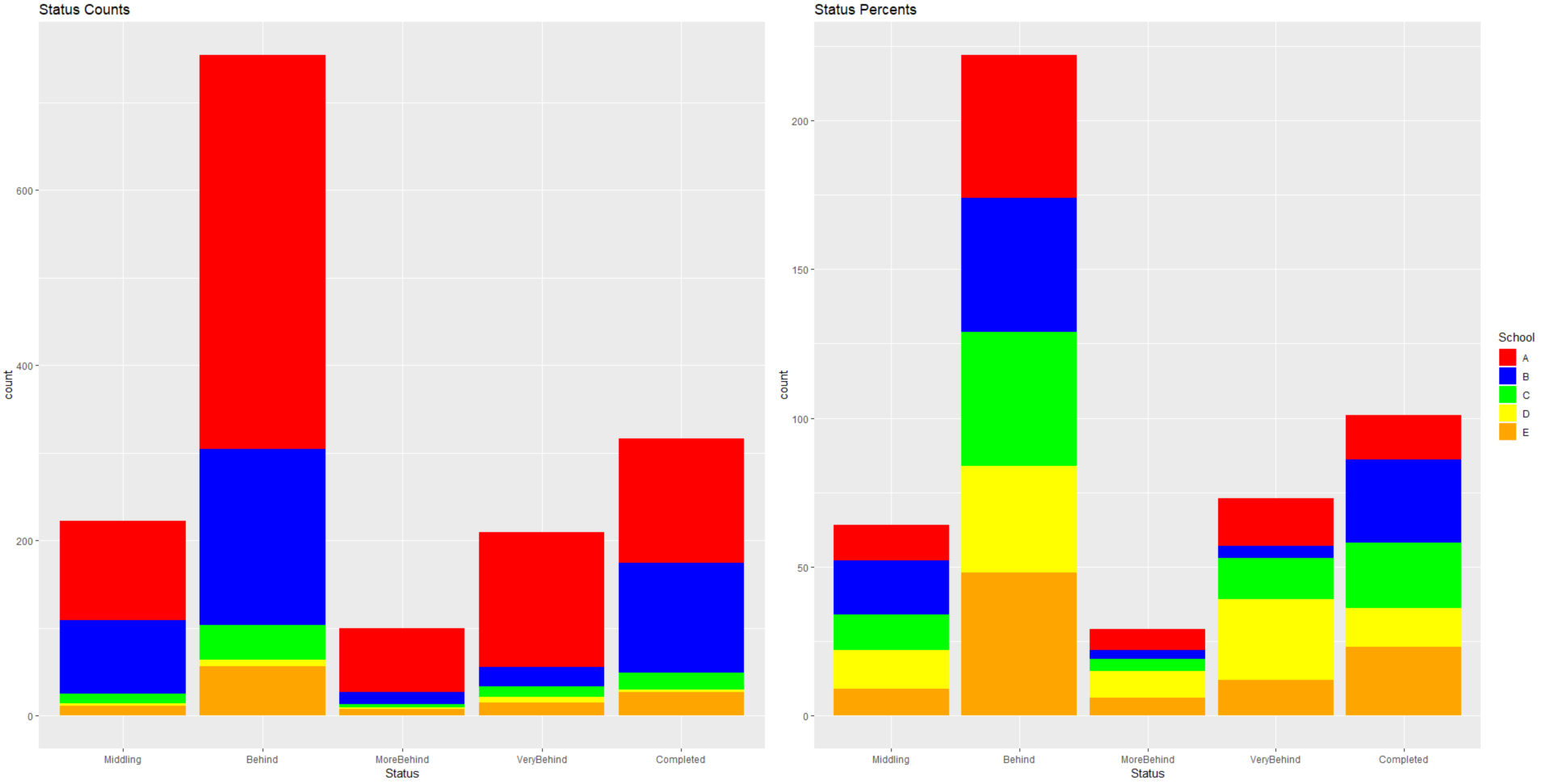


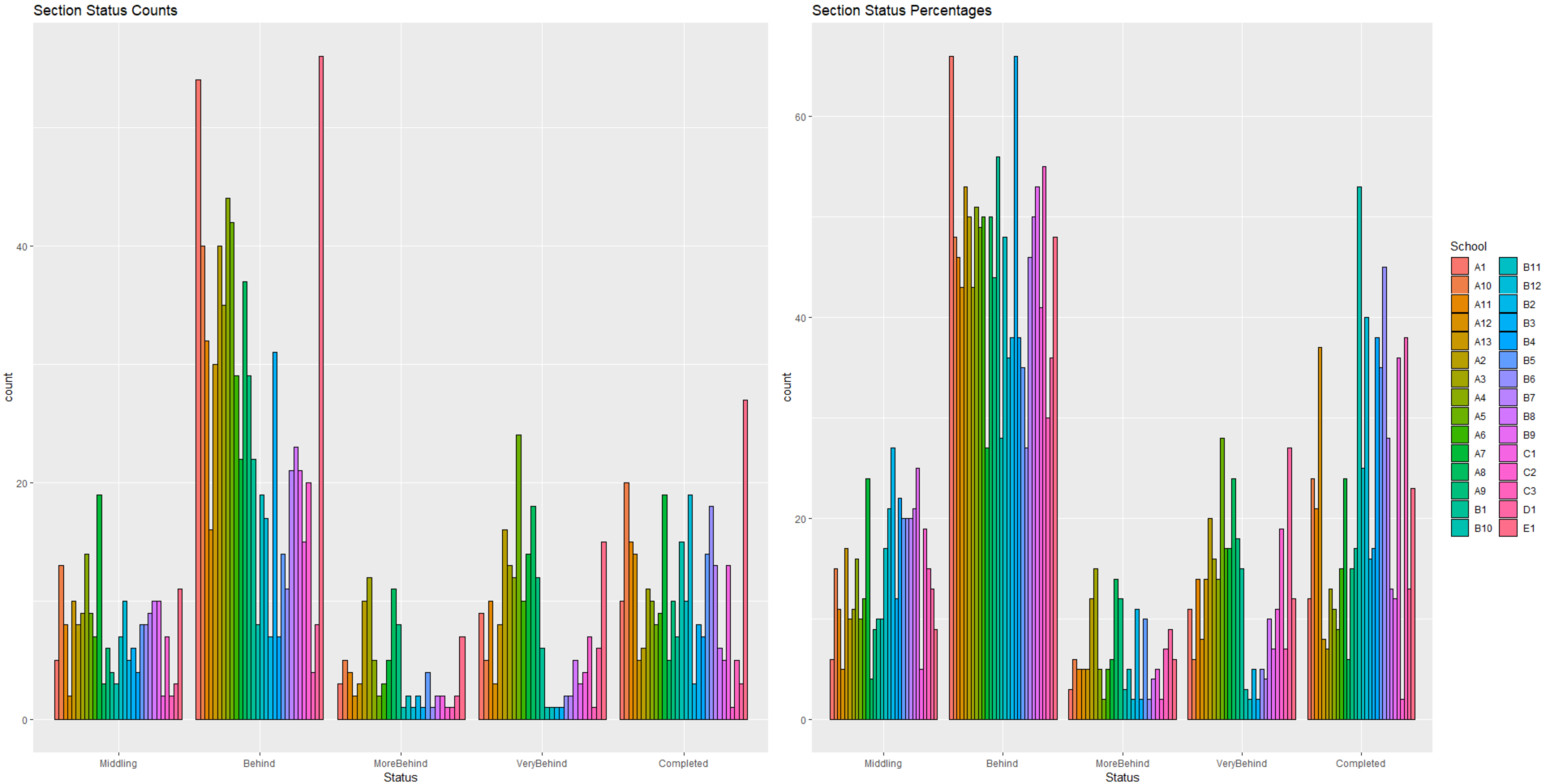
It is also important to note that this data doesn’t include any student scores on assessments (homework, tests, projects) which could be used to better assess the progress of each student, each class, and each school. There is also no note of the grade level of the math program. If this math program is at an 8th grade level, it is possible to have students from 7th, 8th, and 9th grade taking the same material across different schools, or within the different sections.

Each column was plotted against each other and no correlation was found.

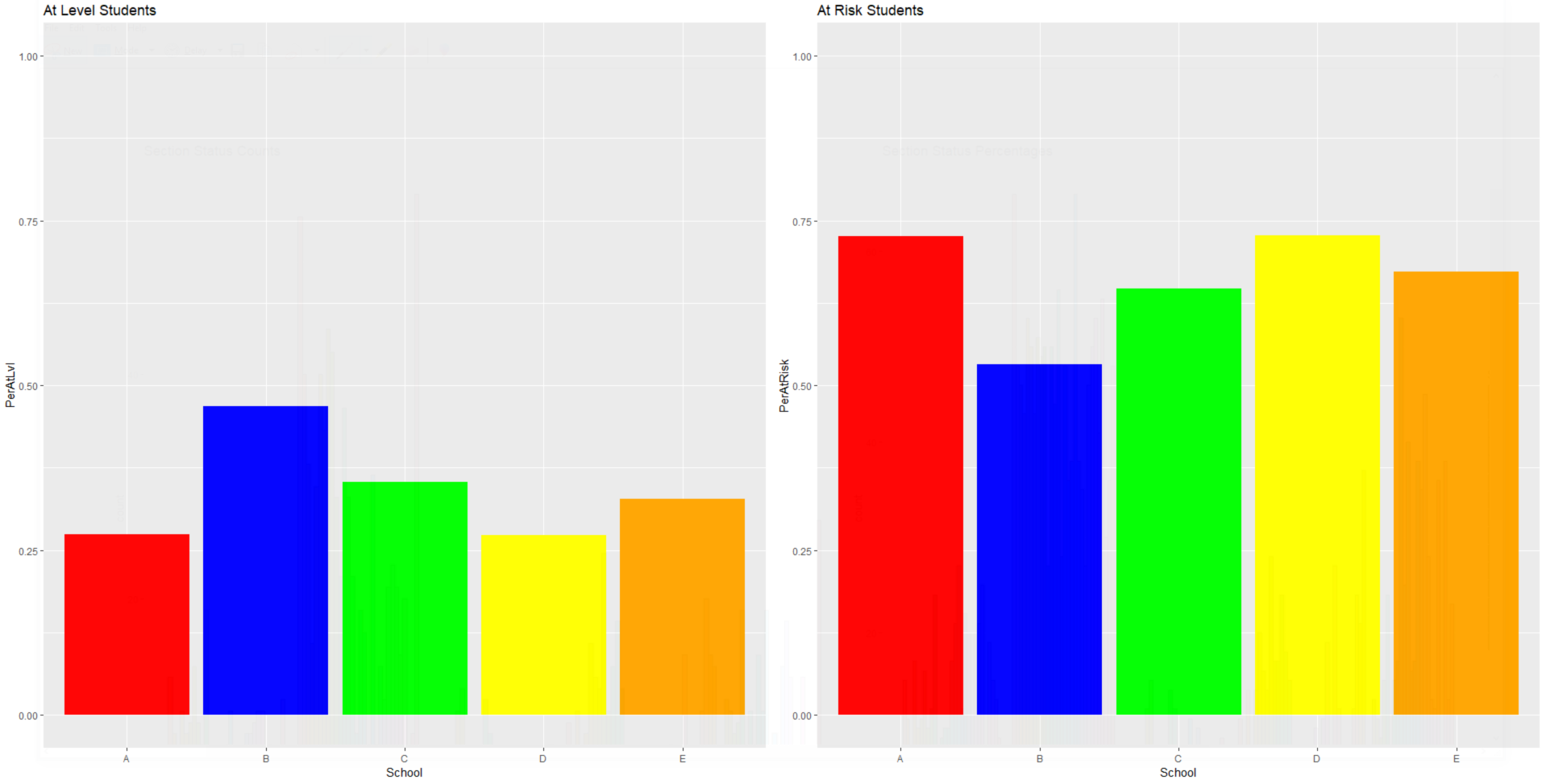


The populations of each school varied enough that visualizing the total number of students within each category for each school did not give a clear representation of the math program or the school’s performance. Transforming the data so that the groups were a percentage of the students in the program and pivoting the columns providing the count of student status so that the status is a factor allowed for plotting all the school data and status. This provides a good way to look at the disparity of the population and the performance of the students per school. The same technique was applied to the section of each school to see if there was a something to be found by examining the students in smaller segments.





Student performance is often broken up into similar categories based on year end test scores. Reorganizing the data by combining the categories to show whether a student is on level (Very Ahead, Completed, Middling) vs at risk (Behind, More Behind, Very Behind) is a good way to examine overall progress of the math program at the school. It can be seen here that the majority of the students (percent) are behind.



**Results**

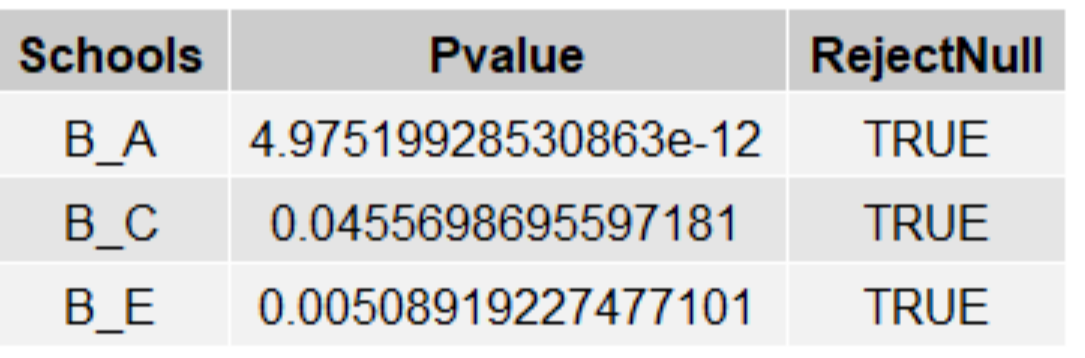
Looking at measures of center, there isn’t enough difference between the median and mean to provide any significant skewness. This may not be an accurate measure of center, however, as there are no assessments assigned to the students.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| VeryAhead | Middling | Behind | MoreBehind | VeryBehind | Completed |
| Min. :0 | Min. : 2.00 | Min. : 4.00 | Min. : 0.000 | Min. : 0.000 | Min. : 1.00 |
| 1st Qu.:0 | 1st Qu.: 4.25 | 1st Qu.:15.25 | 1st Qu.: 1.000 | 1st Qu.: 1.250 | 1st Qu.: 6.00 |
| Median :0 | Median : 7.50 | Median :22.00 | Median : 2.000 | Median : 5.500 | Median :10.00 |
| Mean :0 | Mean : 7.40 | Mean :25.13 | Mean : 3.333 | Mean : 6.967 | Mean :10.53 |
| 3rd Qu.:0 | 3rd Qu.: 9.75 | 3rd Qu.:34.25 | 3rd Qu.: 4.750 | 3rd Qu.:11.500 | 3rd Qu.:14.00 |
| Max. :0 | Max. :19.00 | Max. :56.00 | Max. :12.000 | Max. :24.000 | Max. :27.00 |

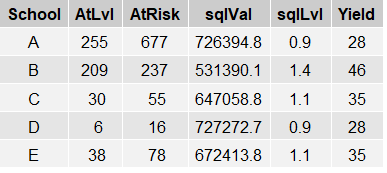
Due to the differences in sample size, a two tailed hypothesis test was performed comparing the school’s populations with each other. The null hypothesis that the both population means were equal was measured to the alternative hypothesis that the population means were different. If the null hypothesis could be rejected, that would imply that something in the school itself, outside the curriculum (teachers, students, parents, method of instruction) was affecting the performance of the students.

In all but three cases, the result of the test was an inability to reject the null hypothesis. The school that was in the test where the null hypothesis could be rejected in favor of the alternative was School B.

* School B vs School A
* School B vs School C
* School B vs School E



Sigma Quality Level was assessed on each school to determine the efficiency of the process. A student falling behind would be a defect. Each student currently behind was treated as a defect against the entire population of the school. Calculating for each school shows they are all performing poorly. No school was performing with a yield above 50%, meaning less than half of the students are not expected to complete the math program. Examining the SQL of the sections revealed that only 5 of the 9 sections in School B and 1 section in school C had a yield above 50%.

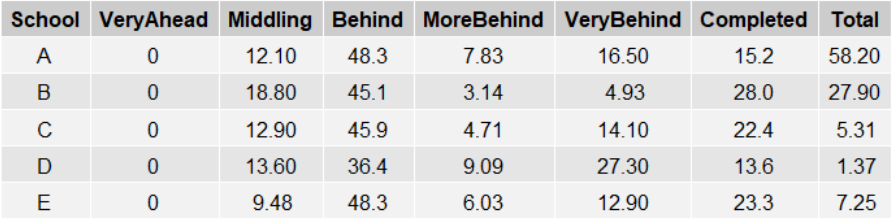


**Conclusion**

The math program

The math program implemented in these 5 schools is ineffective. If the measure of a curriculum is how effective it teaches students by showing them new skills and correcting areas of opportunity, this program isn’t accomplishing those goals. More than half of the students enrolled are currently behind. In only 6 of the 30 sections, with 5 being at School B, are more than 50% of the students on pace to complete the program. This can mean the program is too advanced for some of the students enrolled. There is no data on the students themselves to assess their academic level to make a determination. It is possible, however, that this one program is being used to meet the needs of an entire population that could benefit from multiple curriculums to meet their needs. Students in the ‘Completed’ segment may benefit from a more advanced math course, while students in the behind segments may benefit from remedial math courses. It is very peculiar that students are either ‘Middling’ or ‘Completed’, but none are ‘Very Ahead’. Is this an error in the data? If not, an examination of the final lessons may be warranted, they may not be providing adequate challenge for students.

School B stands out from the others. While its process is performing poorly, it does have the best process, including 5 sections that are outperforming the other schools. It has the highest percentage of students that have completed the course and on pace, and the lowest percentages of students who are behind. It may be of benefit to visit School B and determine what is different about this school that is producing much better results.



Much more analysis would be possible if this data included the school’s demographics, geography, economics, test scores, and grade levels enrolled in the math course. This has caused a few problems. First, it is not appropriate to compare school B with the other schools, but there isn’t enough data to be able to determine what is separating it from the rest. Second, although the length of the course has been given, there is no information about what determines a student moving to the next lesson. Is there an assessment that needs to be passed? If so, what is the passing score? With this type of data, it would be possible to examine how close students are in moving to the next lesson, and be able to examine their progress through the entire math curriculum. Less than 50% of students are on pace to complete the program, but what are the lessons causing the most difficulty? This information would allow an analysis on the course, which is necessary, since it isn’t performing well.