**INTRODUCTION**

In a survey distributed to 5 schools teaching the same math course, results were collected assessing how far or behind students were with their coursework 75% of the way through the course, totaling in 35 lessons. The survey was distributed to see how effective each class’ teaching methods were in regard to student attention and pacing.

As the survey did not collect additional factors that may play into how far or behind a given student is, the results of the following analysis are limited only to student performance within each class.

This survey should provide valuable insights into which classes are utilizing the most effective teaching methods, leading to classroom improvements for further iterations of the course.

**ANALYSIS AND MODELS**

**About the Data**

The survey data set provided contains 30 observations of 8 variables. Each observation corresponds with an individual class, while the variables provide information of which school the class is offered in, the section the class occupies, along with the various measures of student completion.

The completion measures as provided by the surveyor are as follows:

• very ahead (more than 5 lessons ahead)

• middling (5 lessons ahead to 0 lessons ahead)

• behind (1 to 5 lessons behind)

• more behind (6 to 10 lessons behind)

• very behind (more than 10 lessons behind)

• completed (finished with the course)

Upon importing the dataset to the analysis program (RStudio), it is observed that the student counts were in integer format. These observations were converted to numeric as further calculations are needed. The section number also arrived as an integer, but as no calculations are required it was changed to an ordinal data type. Column names were cleaned up to improve legibility (e.g. HW2data$More.Behind..6.10 becomes HW2data$More\_Behind). No missing values were observed. See following for cleaned dataset:

**A screenshot of a cell phone

Description automatically generated**

Figure 1.1

School and Section were merged into a new column (School\_Section).

Upon this level, it was observed that the range of class attendance varied by a large margin (range of 103 students). Due to this discrepancy, the data for completion levels was normalized by converting to a percentage value. See following for resulting dataset (were the prefix “p” signifies percentage values):

A screenshot of a cell phone

Description automatically generated

Figure 1.2

Looking deeper, it is observed that not all of the schools offer the same number of sections for this course (for example School A has 13 sections, while School E only has 1 section). Due to this, the data was manipulated into a new data frame where each observation is one school, and each value is the mean percentage of students taking the course who fall under each category. See the following for the resulting table:

A picture containing white, cabinet, light, display

Description automatically generated

**Models**

The models utilized were visualization techniques, with the intention of gaining a better understanding of the data.

A screenshot of a cell phone

Description automatically generated

The first visualization was a scatter chart displaying how each School Section compared to each other in regard to Number of Students. Number of Students in each Completion Level is differentiated by color, while the schools are shown by usage of different shapes. As class size varies quite a bit, the same visualization was made using percentage values:

A screenshot of a cell phone

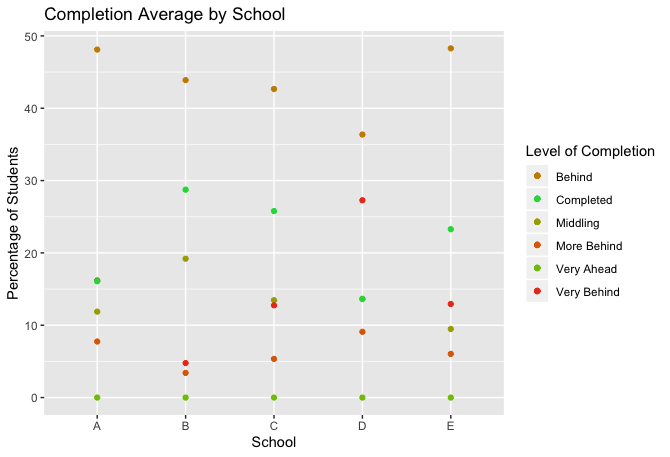
Description automatically generated

This graphic gives a better idea of how each class section might compare against one another.

A picture containing screenshot

Description automatically generated

Going further, this visualization was made to give readers a clearer understanding of how sections competed against each other within an individual school. The schools each have a dedicate chart, with colors still signifying the percentage level of completion.



Lastly, a scatterplot was generated to give a distilled look of how schools compete directly against one another. The sections were averaged together by completion level so only the average values are presented.

**RESULTS**

Based on initial assessments, 35% of students overall are “on schedule” with their coursework, meaning that they are not behind with any of their assignments (Middling, Very Ahead, Completed). Of this number, it appears that School B has the highest percentage of on schedule students (47.9%), while School D has the lowest percentage of on schedule students (27.2%). School B offers 12 sections of the course with an average attendance of 37 students, while School D offers 1 section with an attendance of 22 students.

**CONCLUSIONS**

With respect to the above, a student’s likelihood to be on time with assignments does not appear to have a strong relationship with class size.

School B’s course sections have a higher level of variability than the other school with high section offerings. This might suggest that the professors offering these sections have differing teaching styles, with different student returns. However, this is unable to be confirmed with this dataset, due to its limitations.

It is interesting to note how no students were reported as being “Very Ahead” with their coursework. This could be due to the fact that the survey was given 26 lessons into the 35-lesson term, so it is less likely a student would be more than 5 lessons ahead with only 9 lessons left in the term.