Name: Jie Wang

Email: [JWANG326@Syr.edu](mailto:JWANG326@Syr.edu)

Assignment: HW2

Topic: Math course progress

Models/methods:

Date: Jun, 16,2022

# Explore Math Class Progress

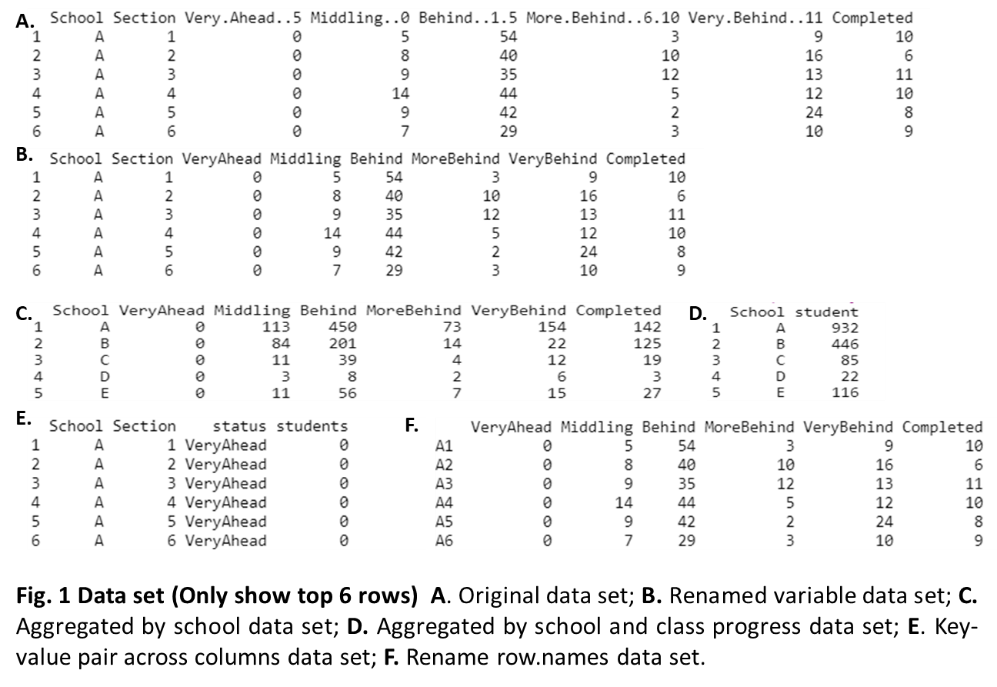
# Introduction

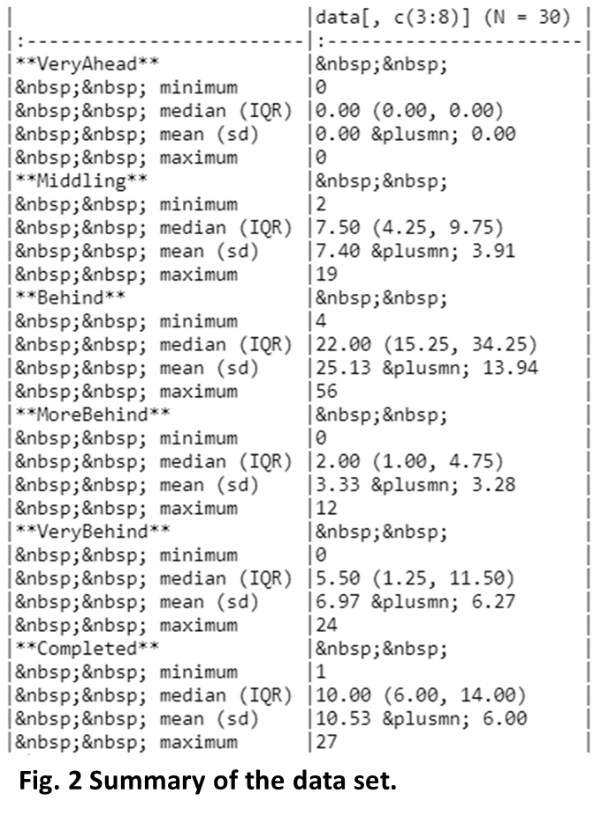
Mathematics is a very important subject; it affects every aspect of our lives. It is a great tool to understand the universe and nature. Schools start teaching math at very early stage, while students’ learning progress may vary due to individua issues or school issues, such as understanding capabilities, attendance, and class schedule.

Learning progress report is valuable for both teachers and students in many ways and considered as a material to evaluate their teaching/learning progress, and assist make a better plan for next steps.

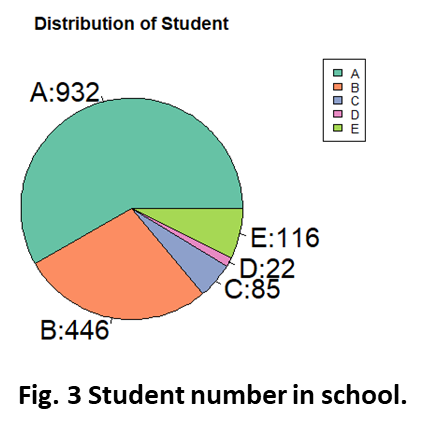
This project is focusing on exploring the math class processing pattern in different schools, including A, B, C, D, and E. These five schools are implementing the same math course this semester. Each school contain vary sections, and total 30 sections. This semester is about ¾ of the way through.

# Analysis

Data preparation and cleaning: original dataset contains variables ‘School’, ‘Section’, ’Vary Ahead + 5’, ‘Middling +0’, ‘Behind -1-5’, ‘More behind -6-10’, ‘Very Behind -11’, and ‘Completed’. ‘School’ and ‘Section’ are nominal variables, and rest of the variables are numeric variables (**Fig. 1A**). Following methods were used to prepare and clean the data: **1)** renamethevariable names to simply the data set (**Fig. 1B**). **2)** is.na () function: To check whether any missing data in the data set, is. na() function was applied. **3)** Aggregation: combine rows within each school to gain total student numbers for each class progress bin (**Fig. 1C**), combine rows and columns to gain student number (**Fig. 1D**). **4)** gather a key-value pair: Using gather () function to generate a key-value pair across multiple columns for ggplot2 data visualization (**Fig. 1E**). **5)** Rename row names: New names were generated by combining column1 (School) and column2 (Section), this data set (**Fig. 1F**) was used for clustering and PCA analysis.

Data exploration: to explore the data, R summary function were used. This data would provide data distribution (**Fig. 2**).

Methods: “tidyverse”, “qwraps2”, “RColorBrewer”, “tidyr”, “forcats”, “ggplot2”, “ggdendro”, “dendextend”, “gmodels”, “ggthemes”, and “ggpubr” libraries and base R were used for analysis and data visualization. ggplot2 library will be used for data visualization. **1)** pie() function in base R was applied to visualize the the overall student number in different schools and in different class progress. **2)** ggplot () function in ggplot2 was used to visualize students number in different class process in school A, B, C, D and E. **3)** Pie chart () will be applied for showing student math class process in overall. **4)** hclust () function in dendextend, which creates dendrogram, was used to cluster the students in 30 sections to explore the class progress similarities between them. **5)** ggscatter () function was used to calculate principal component analysis (PCA) and plot PCA on the data in 30 sections.

Analysis goals: The goal of this exploration is to profile student distribution, their class process status in each of five school, and dig out the which school or section perform well on the student learning progress.

# Results

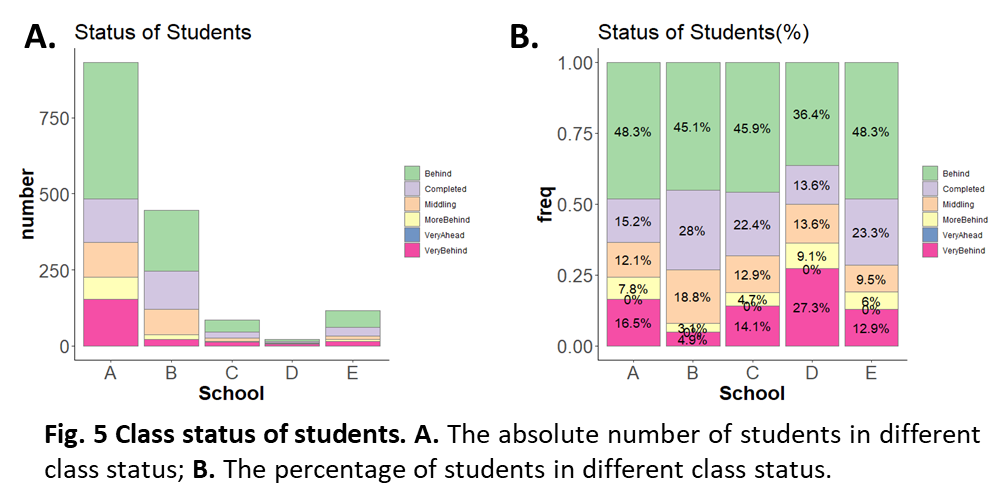
## Student number in each of school

To explore how many students in each of school, a pie graph was generated using base R function. According to the **Fig.3**, schools A and B have a very high number of students, which contain 932 and 446 students, respectively. While school D has least students taking math class, which only contain 22 students.

## Percentage of student in class progress

**Fig. 4** shows the percentage of students in each of class progress status. According to this result, 0 % of students are more than 5 lessons ahead of class (Very Ahead), 13.9% of students are 5~0 lessons ahead of the class (Middling), 47.1% of students are about 1~5 lessons behind of the class (behind), 6.2% of students are 6~10 lessons behind of the class (more behind), 13.1% of students are very behind, which are more than 10 lessons behind, and 19.7% of students finished with the course (completed).These data indicate that majority of the students are behind of the class, and only 33.6% student complete or middling the class.

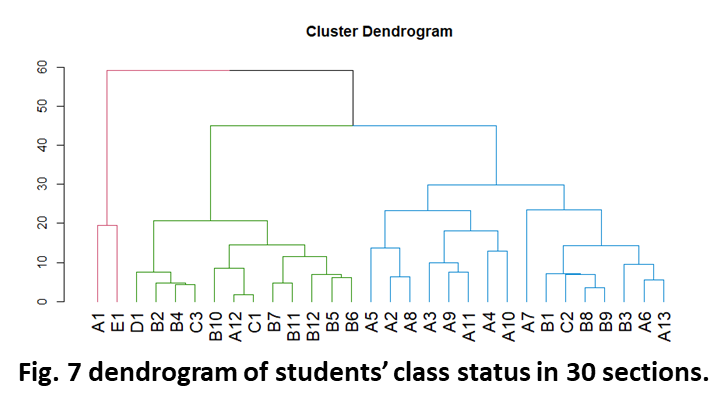
## Class status of students in each of school

**Fig. 5A-B** show the class status of students in each of school (A, B, C, D, and E). Based on **Fig. 5A**, the overall student number is much higher in school A and B, while the students in School D is least, this data is consistent with the observation in **Fig. 3A**. More precisely, School A shows more students are very behind (more than 10 lessons behind, shows in pink), behind (1~5 lessons behind of the class, shows in green), and more behind (5~10 lessons behind of the class, shows in yellow) of the course. **Fig. 5B** shows the percentage of students in different class status in each of school. Students in School B showed a highest frequency of student completed the course (28%, shows in purple) and middling the course (18.8%, shows in orange), and lowest frequency of students more behind (3.1%, shows in yellow) or Very Behind (4.9%, shows in pink) the course as compared with other schools, these data indicate that School B performed very well regarding the class progress. In contrary, School D shows the highest frequency of students who are more behind (9.1%, shows in yellow) and students who are very behind (27.3%, shows in pink) as compared with that in other schools, indicating that School D is performing very poorly.

## Class status of students across each school

Furthermore, Boxpots (**Fig. 6**) showed that only one math class section in school D and E, and three sections in school C. There are more sections in school A and B, while sections in school A shows a big variation in Behind and very behind categories. Students number in school E is highest in class completed and middling, as compared with mean values of other schools. School B shows the lowest mean value of student number in class progress of more behind, and very behind.

## Class status clustering pattern across 30 sections

To explore the school performance similarities between different schools and sections, a dendrogram with three major branches was generated based on students’ class status (**Fig. 7)**. The first branch showed that section 1 in school A showed a very similar students’ class status to that in school E (section 1). Second branch colored in green showed that most of sections in school B (expect section 1, 3, 8, and 9) are grouped together, and also shared the similarity with school C (section 1 and 3), school D, and section 12 in school A. At last, most of sections in school A and section 1,3,8,9 in school B showed a similar pattern, and more precisely, section 2,3,4,5,8,9, and 11 in school A performed very similar regarding the class progress.

## Principal Component analysis (PCA) analysis on class status in each of section in school

Given multiple variables were used to evaluate each section performance on class progression, PCA was then applied to reduce the dimensions. PC1 and PC2 captured most variations. As shown in **Fig. 8** sections in different school showed a similar class progress patten were group together, this data is very consistent to the observation mentioned in dendrogram (**Fig. 7**). Even though groups were not separated clearly, most of sections in school B were more grouped together, and most of sections in school A were grouped together.

Dot size in the PCA plot represent the ratio of student who completed or middling the math class to student who were behind (including behind, more behind, and very behind, the bigger dots mean the section perform well on the class progress, while the smaller dots indicated that the section performed poorly. Based on this, it is clearly showed that school B, especially section 4,5,6,10 and 12, performed very well and showed a great ratio of class finish or ahead. While school A, especially section 1,2,5, and 8, performed very poorly and showed a lower ratio of class finish or ahead.

# Conclusions

The aim of this project is to determine students’ learning progress in math and plan the next steps towards achieving the learning goals. This study indicated that school A contain highest number of students, while school D contain lowest number of students. In total, there around 70% students are lagging behind, only around 30% complete or ahead of class.

More precisely, school A and school B hold 13 and 12 sections, respectively. Most sections in school B (such as sections 4,5,6,10 and 12) performed very well, they show a high ratio of students who completed or ahead of the class, while most sections in school A (such as sections 1,2,5, and 8) performed poorly and have a big variation between different sections. These data indicate that school teaching quality may involve in student progress.

Both teachers and students could benefit from this data, it allows teachers understand students learning progress, and adjust their teaching strategies in following weeks to promote student learning. For the students, this data report can reflect their learning and learning strategies, and with that, they can make adjustment to make better progress in the class.