John Fields
Dr. Ami Gates
IST707 - Assignment #2
7/20/19

Introduction

Considering the importance of education, it is useful to have a framework which provides insights on how to improve school systems holistically. Stanford University has identified nine "grand challenges" for education and learning that provide a good starting point for understanding these priorities:

1. Transformative Technologies
2. Better Systems of Education
3. A Worthwhile Education
4. Science of Learning
5. The Promise of Every Child
6. Research and Practice Exchange
7. A Data Revolution
8. Empowered Educators
9. Identities and Participation

The purpose of this paper is to look at the information from a small sample of schools to analyze the data for a math course to understand trends and provide recommendations on potential changes. This information can be used to inform administrators, educators, parents, students, and government officials on potential priorities to improve or correct issues.

As comparisons are often made between schools, it is important to keep in mind the bigger picture and not fall into the trap of measuring and optimizing one area while negatively impacting other areas. As the small data set for this assignment is analyzed, conclusions will be developed with the understanding that much more data and analysis would be needed prior to implementing any recommendations which could have a big impact on the current and future students in these schools.

Analysis and Models

About the Data

The data for this assignment includes information from 5 schools (A, B, C, D and E) on a math course with 35 lessons and 30 sections total between all the schools. The data provided provides a snapshot of student progress for approximately 3/4 of the way through the semester. Table 1 below provides a summary of each variable that was provided in the data set.

Variable	Definition	Range of Values	Type
School	School identifier	A,B,C,D,E	Categorical, Nominal
Section	Section identifier	1-13	Categorical, Nominal
Very Ahead	Students 5 or more lessons ahead	0	Quantitative, Interval
Average	Students 0 to 5 lessons ahead	2-19	Quantitative, Interval
Behind	Students 1 to 5 lessons behind	4-56	Quantitative, Interval
More Behind	Students 6 to 10 lessons behind	0-12	Quantitative, Interval
Very Behind	Students 10 lessons behind	0-24	Quantitative, Interval
Completed	Students who finished all lessons	1-27	Quantitative, Interval
School Size	Schools grouped by size	Small, Large	Categorical, Ordinal

Table 1: School Data Variables by Type

The first step in the exploratory data analysis (EDA) was to understand the relationship between each variable and the analysis goal. The assumed goal for this assignment would be to have all students in the "Completed" status by the end of the semester and increase the completion rate for future students.

Prior to beginning an analysis of the data, cleaning and prep was needed to remove spaces and change the variable names to more concise titles as shown in Table 1 above. The second step was to remove the "Very Ahead" category since all 0 values likely indicates a data quality issue with this variable. Since this was such a small dataset, minimal manual processing was needed to prepare the data for analysis. Below is a summary of the statistical data for this data set.

School	Section	Average	Behind	MoreBehind	VeryBehind	Completed	SchoolSize
A:13	Min. : 1.00	Min. : 2.00	Min. : 4.00	Min. : 0.000	Min. : 0.000	Min. : 1.00	Large:25
B:12	1st Qu.: 2.25	1st Qu.: 4.25	1st Qu.:15.25	1st Qu.: 1.000	1st Qu.: 1.250	1st Qu.: 6.00	Small: 5
C: 3	Median : 5.50	Median : 7.50	Median :22.00	Median : 2.000	Median : 5.500	Median :10.00	
D: 1	Mean : 5.90	Mean : 7.40	Mean :25.13	Mean : 3.333	Mean : 6.967	Mean :10.53	
E: 1	3rd Qu.: 9.00	3rd Qu.: 9.75	3rd Qu.:34.25	3rd Qu.: 4.750	3rd Qu.:11.500	3rd Qu.:14.00	
	Max. :13.00	Max. :19.00	Max. :56.00	Max. :12.000	Max. :24.000	Max. :27.00	

Table 2: Summary Statistical Data (Section information is not meaningful since these are qualitative variables)

Models (methods used to analyze the data)

After importing the data into the analysis software, several graphs were created to better understand the overall data prior to beginning more detailed analysis. Figure 1 below shows that Schools A and B had many more students than Students in Schools C, D and E.

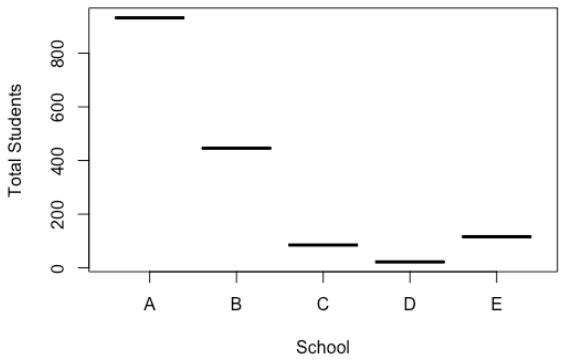


Figure 1: Total Students Per School

After understanding these groupings during the exploratory data analysis (EDA), the decision was made to create a new feature by combining Schools A and B into a "large school" group and combine Schools C, D and E into a "small school" group.

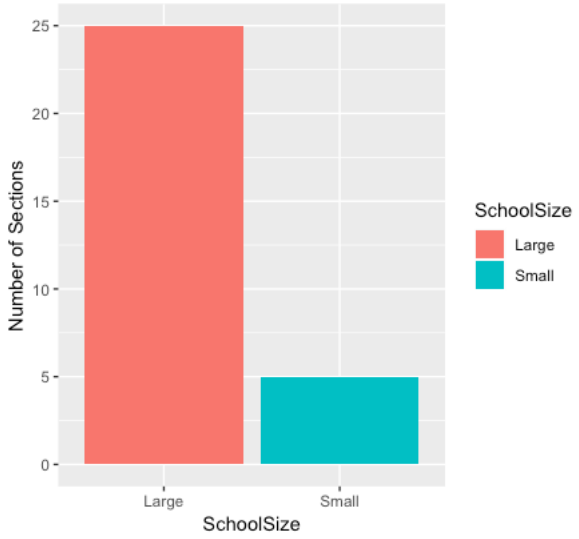


Figure 2: School Size and Number of Sections

After understanding the variations in the total number of students per school, it was also important to understand the number of students who were complete ("Complete") and likely to complete ("Average") compared to those students who are behind ("Behind", "More Behind", "Very Behind"). Figure 2 below provides this comparison. The number of students Behind is 66.4% and those who are Complete or Average (expected to complete by the end of the semester) is 33.6%.

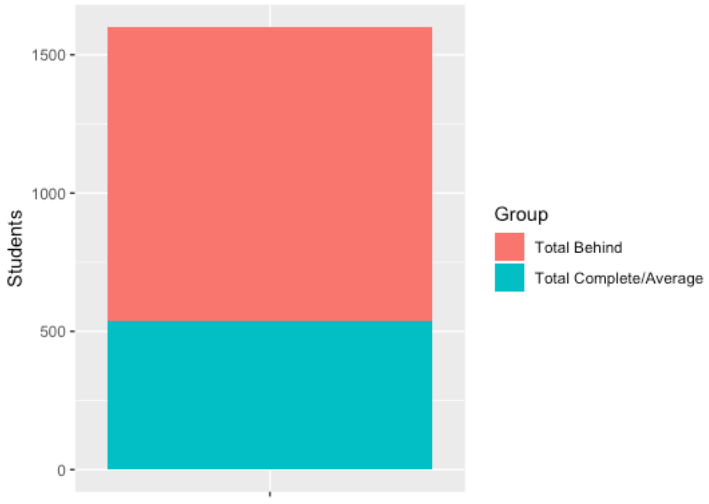


Figure 3: Total Expected Completion Rates

The percent complete by school was then graphed as shown in Figure 4 below to understand the differences in the actual completion rate between each of the schools and the number of students.

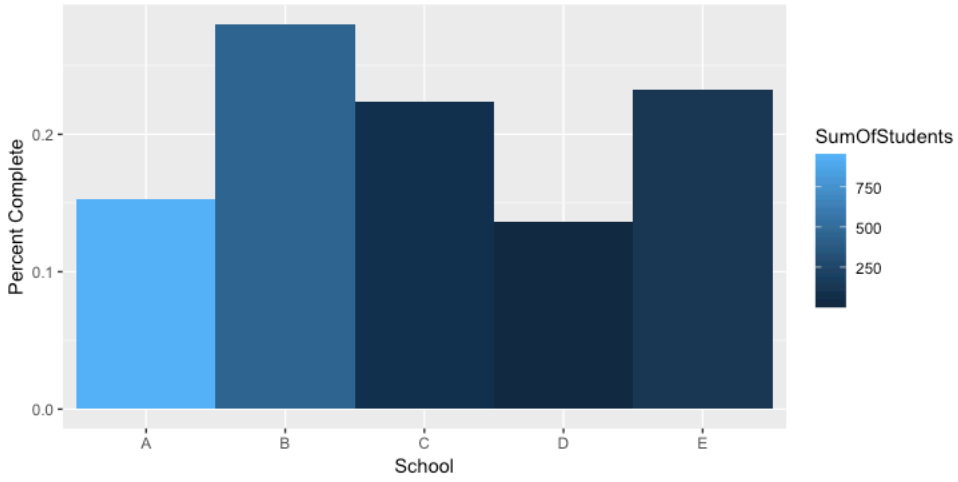
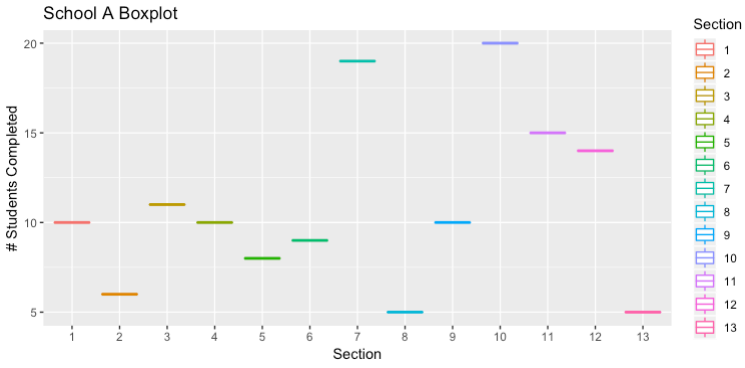
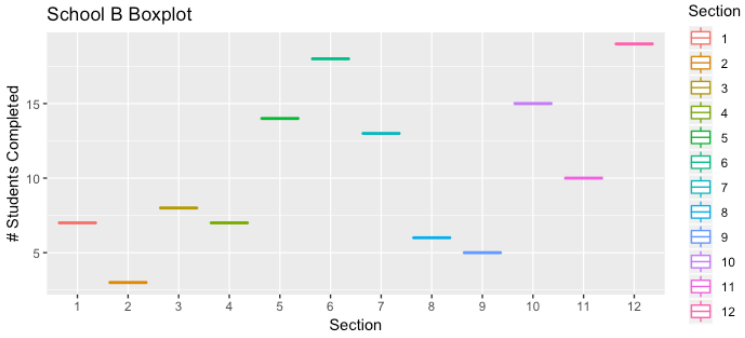
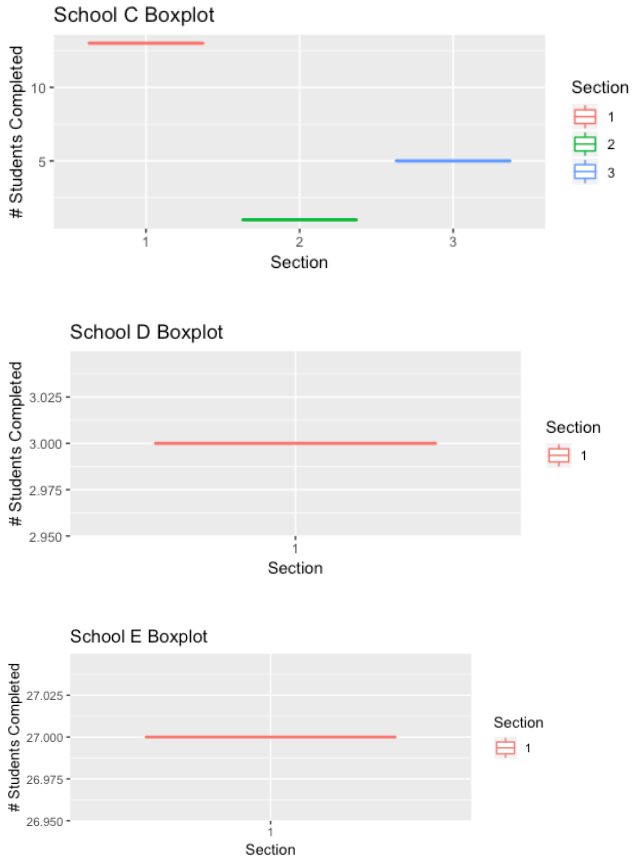


Figure 4: Percent Actual Completion Rate for Each School

Next, to better understand the variance in the actual completion per school, boxplots were created to see the variation in the rates per school by Section.







Figures 5-9: Actual Completion Rates by School

Another factor to be considered for the effectiveness of learning is the class size. It is assumed that each "section" of the data is a different math class. However, as shown in Figure 10 below, there are some anomalies in the data with School E has 116 students in 1 section and School B has 5.7 students per section. School E probably has multiple sections but is only showing 1. If the data for School B is accurate, then the low student per section of 5.7 could be a contributing factor since School B has the highest actual completion rate at 28%. However, the cost to have 5.7 students per section would be very high.

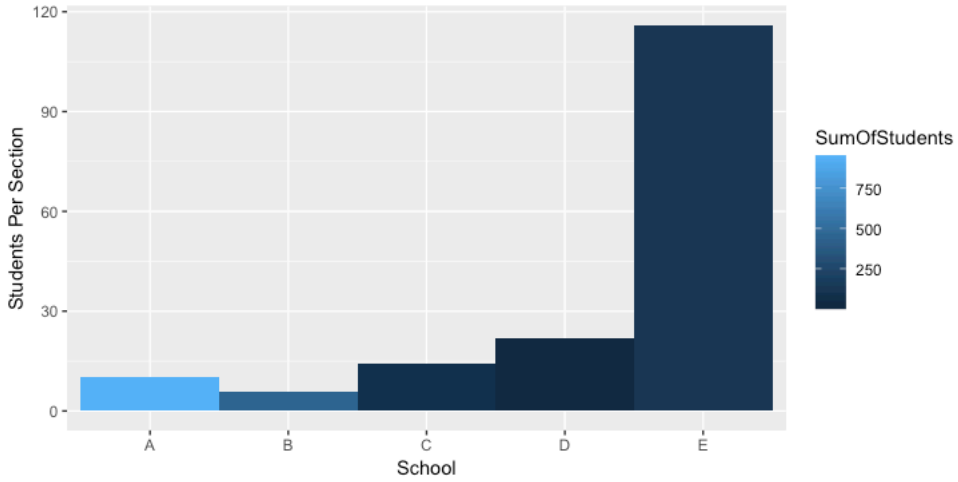


Figure 10: Students Per Section by School

Results

The analysis of the math classes from 5 schools shows that only 1/3 of the students will likely complete the class by the end of the semester. This includes those students who have already completed the class plus students in the "Average" category who are expected to complete the class (0-5 lessons ahead).

Below is a summary of the seven observations that resulted from the analysis of this data:

1. 33.6% of the students from all schools are expected to complete the class (“Complete”/”Average”) and 66.4% of the students are behind (“Behind”, “More Behind”, Very Behind”).

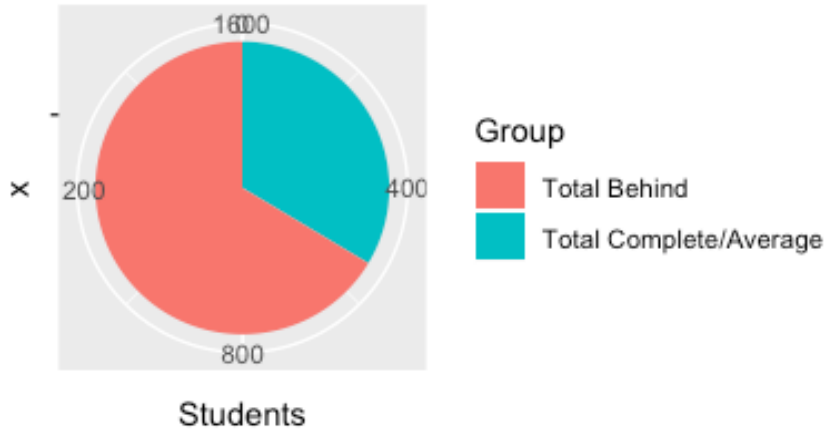


Figure 11: Percentage of Students Expected to Complete (Total Complete/Average) and Not Expected to Complete (Total Behind)

2. School A has the largest number of students (932, 58% of total) and the lowest actual completion rate (15%).
3. The large schools (A & B) and small schools (C, D & E) have significantly different numbers of students and sections.
4. Schools B, C & E have a better actual completion rate (22-28%) compared to schools A & D (13-15%).
5. There is a wide variation in the Actual Completion Rates (Figure 4) within each school from section to section in the large schools.
6. Students in School A - Sections 7/8, School B - Section 6/12, and School C - Section 1 are doing better than other sections at the same school. Unfortunately, there are also several sections that are performing poorly, so investigating School A - Section 2/8/13, School B - Section 2, and School C - Section 2 would be helpful to understand why these sections are below average.
7. The students/teacher ratio is a factor that has been shown to influence the learning success for students. The lower student/section ratios of 5.7 for School B and 14.2 for School C appear to be correlated with improved completion rates for students. However, cost data was not provided for this analysis so this would need to be included before making any changes.

	School A	School B	School C	School D	School E
Students Per Section	10.2	5.7	14.2	22.0	116.0

Table 2: Students Per Section by School

After understanding these groupings during the exploratory data analysis (EDA), the decision was made to create a new feature by combining Schools A and B into a "large school" group and combine Schools C, D and E into a "small school" group. To analyze and draw

conclusions about the small data set for these five schools, it is important to remember that the only information provided is students who are completed vs. not completed.

Conclusions

The data supplied for these five schools indicates a major issue in the math class analyzed since only 1/3 of the students are expected to be complete by the end of the semester. With 2/3 of the students behind, there is concern across all schools since the expected completion for the semester ranges from 27-48% (see Figure 11 below). However, since no information was provided to understand how well students are learning, data should be collected to understand knowledge acquisition to compare with the available data on completion of the class.

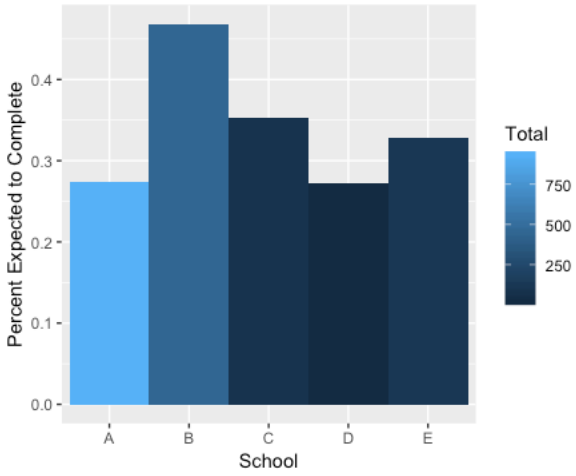


Figure 11: Percent Expected Completion ("Completed", "Average") by School

In addition to the main conclusion above, other key observations and recommendations are summarized below:

1. School A has the greatest issues impacting the largest number of students (932) and the lowest actual completion rate (16%). A different approach for any change is recommended for the large schools (A & B) vs. the small schools (C, D & E) due to the size differences. Looking at the reasons for the success or challenges in these groups could provide insights for how to improve the overall success for all schools.

2. A better understanding of the reasons for the success of high completion sections vs low completion sections (Figures 5-9) may provide insights on how to improve the overall consistency from section to section. Some teachers are performing better than others and it would be helpful to observe the successful teachers to understand what makes them connect with students to achieve higher levels of success.

Administrators, teachers and parents should consider these recommendations as a starting point to understand the current situation and what changes need to be considered to make improvements. There is insufficient information to determine a root cause related to the math curriculum, training for teachers, or other factors. It is also difficult to discern from this data how well students are learning and which of the "grand challenges" listed above need to be addressed. However, an investment of time and resources is recommended to better understand the low completion rates in this math class and address the other challenges described above.

In conclusion, education is a critical factor to ensure that students have the skills and knowledge to lead a productive and healthy life. A quote from David E. Campbell's paper, *What is Education's Impact on Civic and Social Engagement?*, is included to provide a reminder on the importance of education as an engine for economic growth and for boosting the "social capital" of a society.

> "While policy makers widely recognize the fact that education serves as an engine for economic growth through the accumulation of human capital, education is also strongly associated with boosting levels of social capital. Indeed, an important justification for the large expenditures on education within many democratic nations is its social, and not just economic, impact – the benefits an educated electorate brings to civil society."

Works Cited

"Grand Challenges." *Grand Challenges | Stanford Graduate School of Education*, 2019, ed.stanford.edu/vision/educational-grand-challenges.

Campbell, David E. "Symposium on Social Outcomes of Learning." *What Is Education's Impact on Civic and Social Engagement?*, 2006.