Sana Khan

IST 707

Homework 2

Each of 5 schools (A, B, C, D and E) is implementing the same math course this semester, with 35 lessons. There are 30 sections total. The semester is about 3/4 of the way through.

For each section, we record the number of students who are:

• 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)

Background:

This dataset is showing the number of students that fall into one of the statuses for the math course section that their respective school is in. There is a total of 1601 students, but no indication of the grades or locations of the five schools involved. We can also see from the dataset that the number of sections in each school is different among the five schools. We can not infer from the dataset why there are different sections among the schools either, but we do have data on the number of students in each school and each section. This data can help us understand where action needs to be taken.

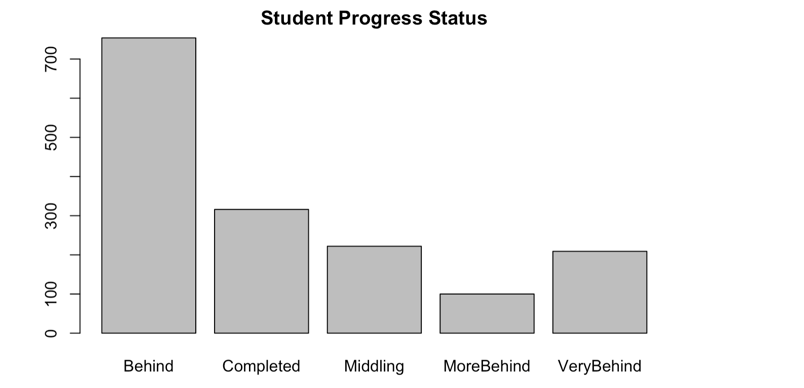
Analysis:  
  
I began by grouping the progress statuses by school and creating a new data frame to understand the data. I also used this table to understand the number of students per school. This also showed that there were not any students in the Very Ahead status so I removed it from the data frame.

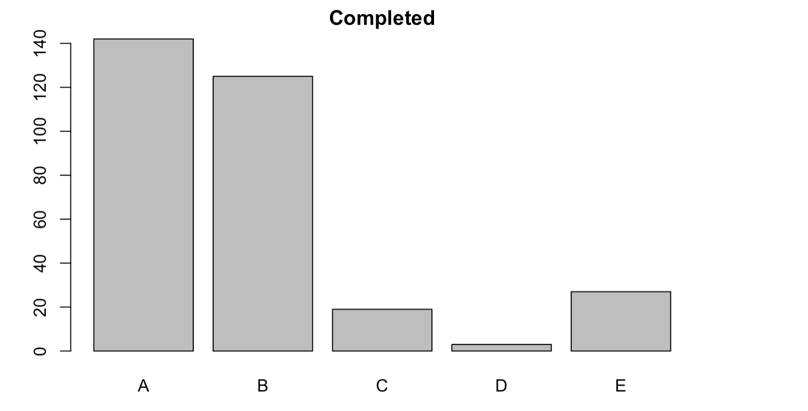
From the data we can see that School A has the most students with 932 students. School A also had the most number of students in the Behind Status, which is 450 students so almost half of the students. School B has 446 students with 201 in the Behind Status which is about 40% of the student population. School C has 85 students with 39 in the Behind Status which is about 45% of the student population. School D has 22 students with 8 in the Behind Status which is 36% of the student population. School E has 116 students with 56 in the Behind status which is 48% of the student population.

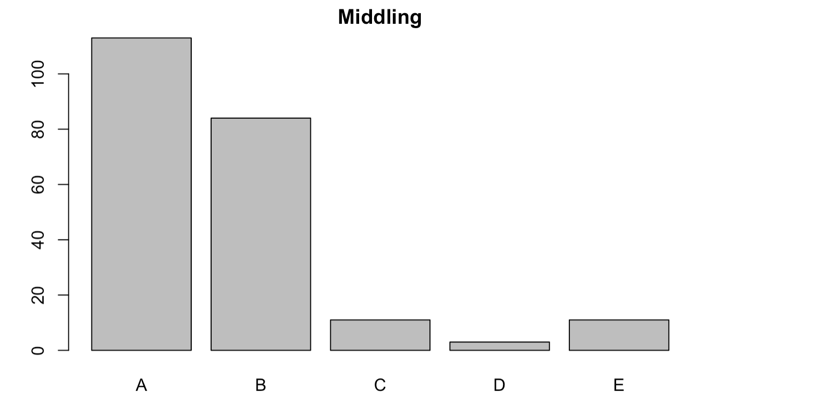
From the results each school had the most number of students in the Behind Status compared to the other statuses. The completed status is second in the number of students, with middling after. If we are to combine the status of Behind, more behind and very behind into one group than all schools have more than 50% of students falling into one of these categories, with schools A and D with more than 70% of students falling into this grouping.

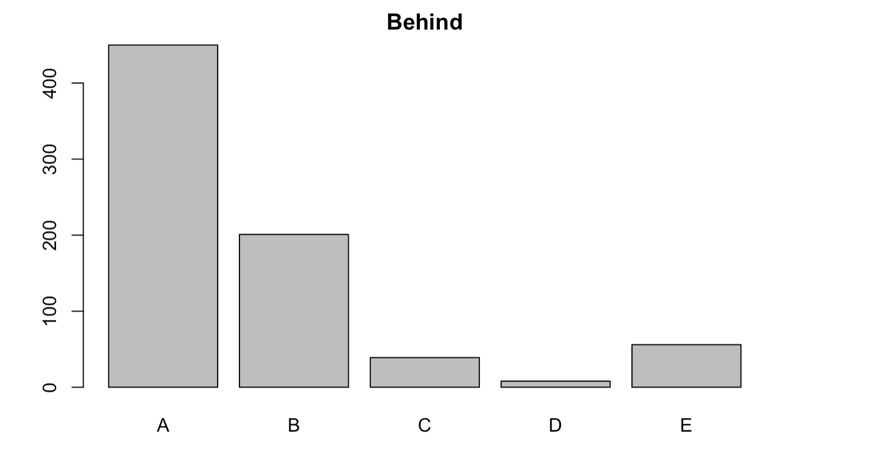
I then grouped the data based on sections to see if there were any trends based on the section. Based on this data it appears that most students are falling behind or greater in section 1.

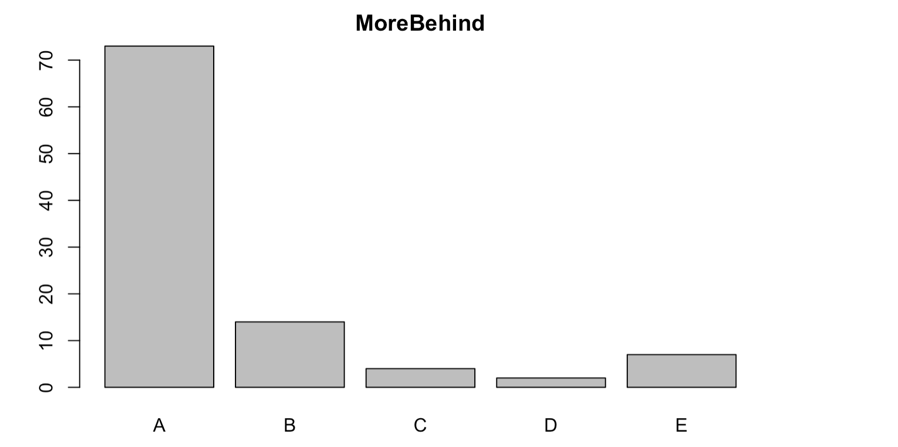
I then graphed each status separately to have a visual representation of the data.

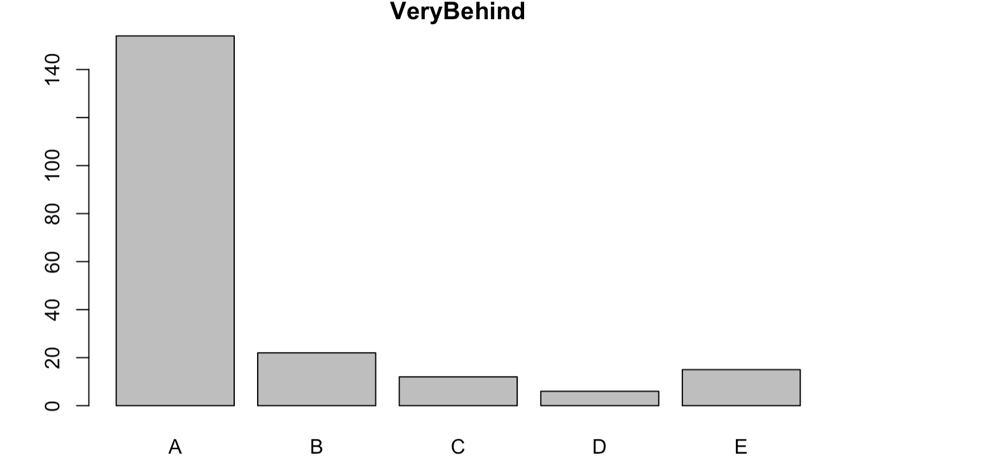












Conclusion:

Based on the data, all of the schools are at least 50% or more behind in completing the sections with school D performing the worst relative to the other schools. It also appears that schools are performing the worst on the first section, which may bleed into the other sections not being completed either. Although school D did perform the worst, it could be an outlier and may need deeper investigation. However, based on the data it does show that smaller schools seem to be performing better than the larger schools. This could be because of resources not being allocated properly, or too many students per teacher ratio. There are also questions about how and when this data was recorded and how a section is defined as well as what constitutes a secion being complete.