Jeff Levesque

Professor Gates

IST 565

Homework #2

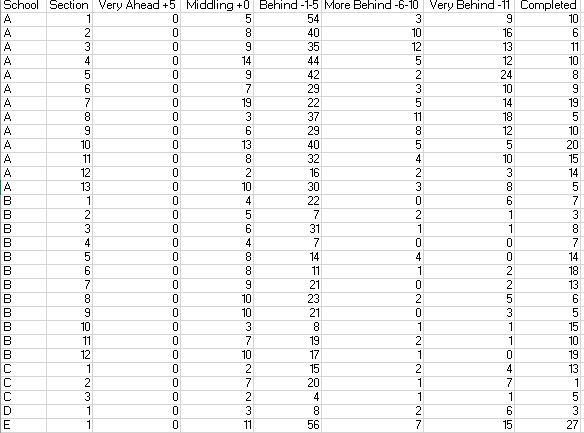
**Introduction**

Often college departments will consider adding or preserving a course in a curriculum. However, the success of the course, in context of continual availability, should be predicated primarily on two measures. First, the demand of students wanting to take the course, and second the successful attrition of students completing the course. As a precursor to measuring the attrition of students, understanding whether students are keeping up with the course is an important understanding.

**Analysis**

Data Preparation

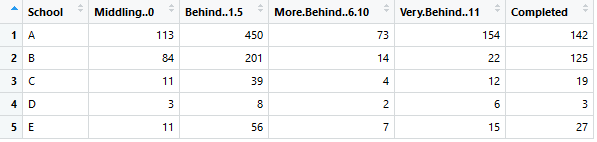
A csv dataset representing students from five anonymized schools, and corresponding sections have been contained in a csv file



The first column represents the anonymized school, categorized alphabetically from the letter A. The second column represents the section within the school. These values were unique, and sequential starting at 1, for each school. Successive columns were positive integers representing the number of students who were accounted for the following categories:

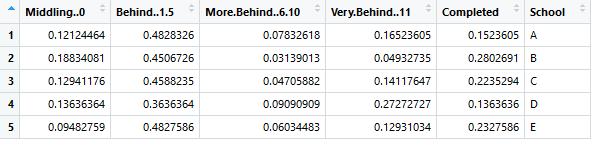
* Very Ahead +5
* Middling +0
* Behind -1-5
* More Behind -6-10
* Very Behind -11
* Completed

Using R, the dataset was loaded into a dataframe. Any columns containing all zero rows, were eliminated. In the given dataset, this eliminated the Very Ahead +5 category, since no student from any school was ahead. Next, to simplify the dataset, the Section column was removed. Then, the dataset was aggregated by the School column. This meant that all rows belonging to a school (i.e. A), was collapsed into a single row, where the corresponding column values were summed, across all possible rows in the given class:

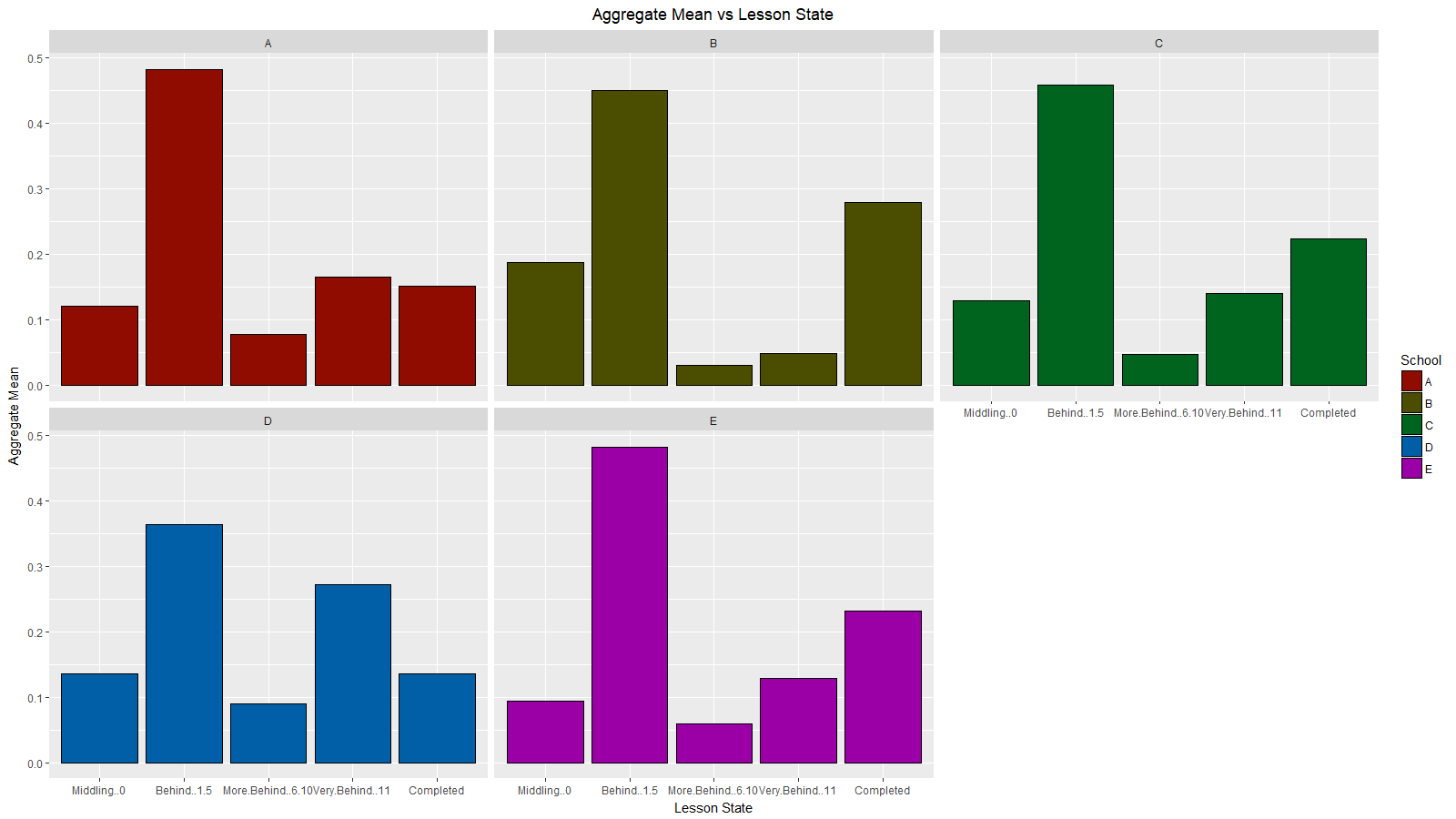


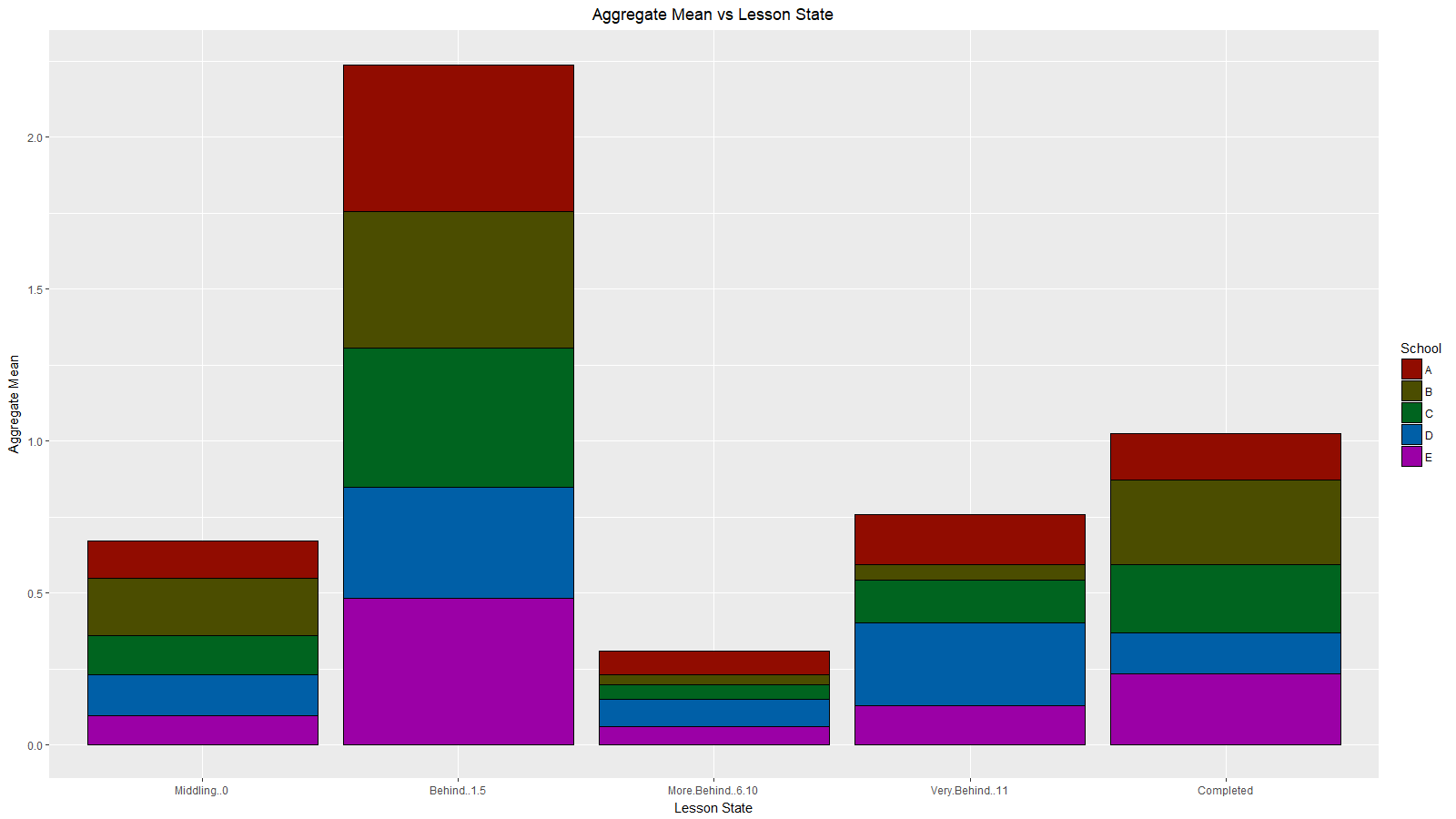
Processing:

After the dataset was adequately prepared, row sums were computed, producing a frequency distribution. However, each cell in the dataframe was divided by the corresponding row sum. This standardization allowed each School, regardless of the number of students, to be compared to another School, based on a distribution ratio:



To generate the corresponding visualization, this dataframe was converted from a wide to long syntax, which allowed correspond ggplot2 bargraphs to be created:





**Results**

The generated visualizations indicate a general pattern. Specifically, most students were generally Behind -1-5, in all schools. The next category Completed was slightly ahead of Very Behind -15, followed by Middling 0, and More Behind -6-10. Some limitations of the generated visualizations may lie with the dataset. For example, when Class D was collapsed into a single row, the corresponding column sums were all single digit values. Other classes contained mostly all double digits, while some had triple digit sums. This may indicate that the data was not fully representative of the School from where it was collected from.

**Conclusions**

The dataset represents students ¾ through the semester. To instructors, and department leads, discovering student generally behind, would be an alarming negative finding. However, it would be beneficial to the department, to obtain a secondary set of results. These results could potentially be the student’s final grade in the course at the end of the semester. Having these values, in addition to the ¾ checkpoint, could allow a model to be generated. Specifically, the checkpoint could attempt to predict the final grade. However, to further on the idea, the arbitrary ¾ checkpoint could potentially be converted to a timeseries data, with regular points representing various class deadlines. This could potentially allow timeseries forecasting. Therefore, instead of only visually noticing at ¾ through the semester, instructors could use forecasts to mitigate students who are following a trend, which could lead to poor final grade performance. By using timeseries data, along with possible IMR (i.e. signal processing), student trends could alarm instructors to be more proactive, to realign students to their full potential.