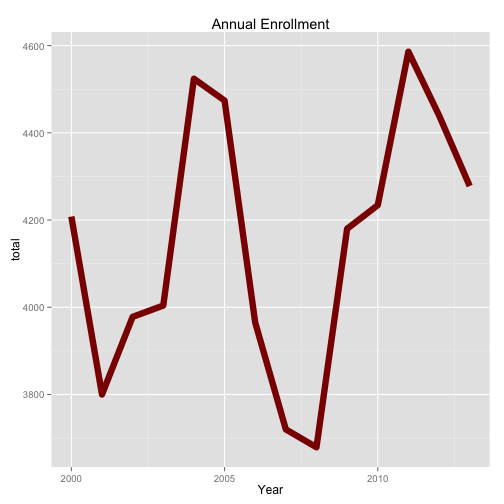
# Department Report

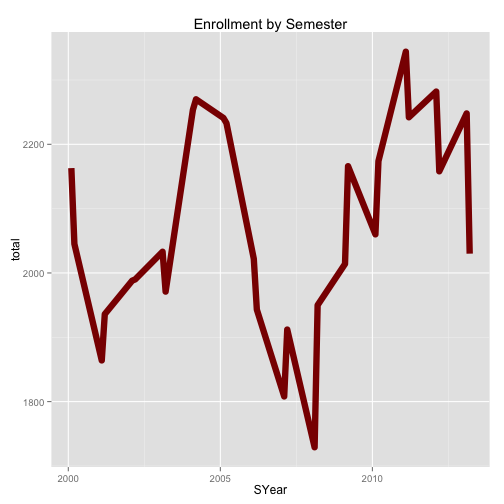
The purpose of this analysis was initially to look at how the addition of online courses has changed the enrollment at branch campuses. A number of other questions were also asked, almost all of which dealt with how to build a class schedule that maximized undergraduate enrollment. This paper will begin with showing some enrollment trends over the past decade and then attempt to address each of the questions that have been asked.

## Enrollment Trends

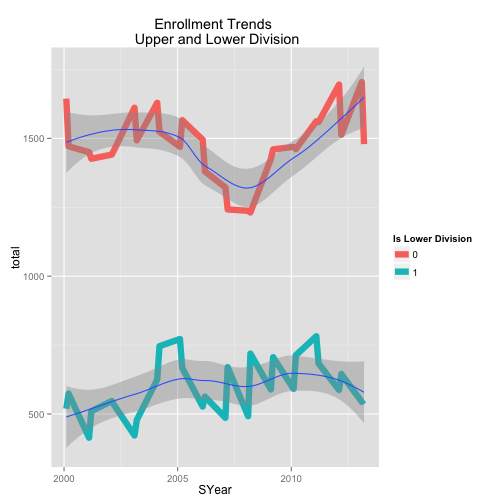
The data for this project covers the time from Spring 2000 to Fall 2013. Summer enrollment was not gathered. In the graph below we see that annual enrollment (calendar year) generally oscillates around 4000 students. The dramatic drop at the end is spring semester 2014.



In this time period the maximum enrollment was in 2011 with an enrollment of 4586. If we look by semester there is much more variation as shown in the graph below. The low point was the Spring semester of 2008 with a total enrollment of 3679.

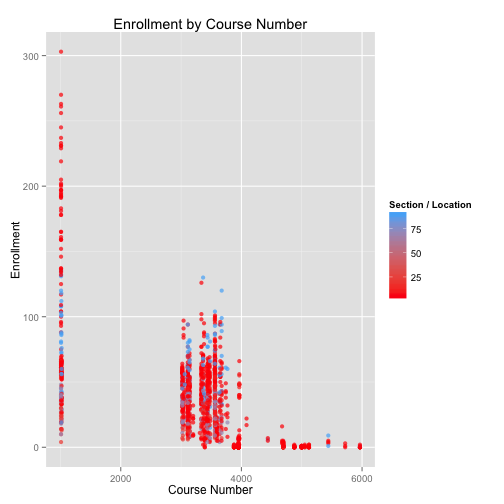


Over the 14 years of data that are available a total of 1514 courses that have been offered and 58070 seats enrolled. There have been 83 unique course numbers. The courses themselves have a very different enrollment patterns and shown in the graph below that shows the difference between upper and lower division courses.



The large drop in enrollment in 2008 appears to corresponds to a decrease in upper division courses while the lower division courses remained relatively stable, as indicated by the trend line.

As we move on to looking at variation in individual courses we can also start to look at branch campus information as with the graph below. The course numbers are arranged on the x-axis. The color adds the additional information of where the course is being held (section numbers). Sections that are red (less than 10) are course held at the main campus while values of 90 (blue) are online. The light red in the middle represent branch campuses.

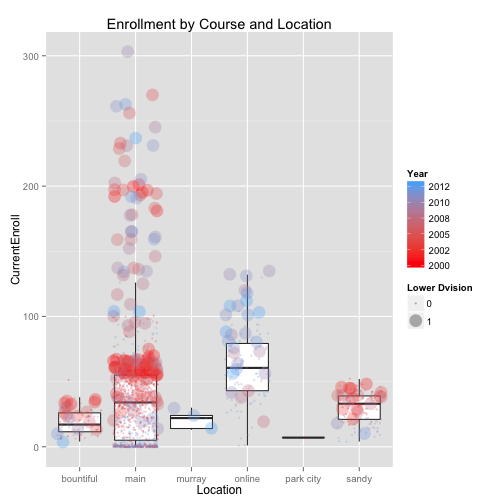


## Campuses

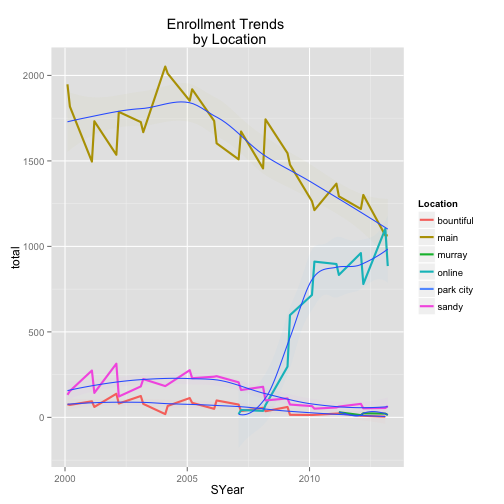
Lower division courses have substantially higher enrollment than any other course. This is primarily driven by SOC 1010 and the highest enrollments occur at the main campus. The main campus generally has the highest enrollment for each course, with the exception of some of the online courses. The table below shows the total enrollment by location. What is notable is that online courses have the second highest number of students enrolled although they only started to be offered in 2006.

|  |  |  |
| --- | --- | --- |
| Location | Enrollment | Number of Course |
| main | 44071 | 1159 |
| online | 8153 | 130 |
| sandy | 4223 | 140 |
| bountiful | 1513 | 79 |
| murray | 103 | 5 |
| park city | 7 | 1 |

If we look at a box plot of the different campuses we see that online courses have a higher mean than any other location. As shown above the highest enrolling courses are lower division courses, and this has remained constant over time. It appears that the higher enrolling courses at Bountiful and Sandy were older, and the current courses have lower enrollment.



The trend lines (in blue, below) show a decline in main campus enrollment prior to the introduction of online courses. The branch campuses appear to have remained fairly steady in their enrollment until the introduction of online courses. Online courses, since their inception, have grown quickly. The rate of growth appears to have leveled off in the past few years.

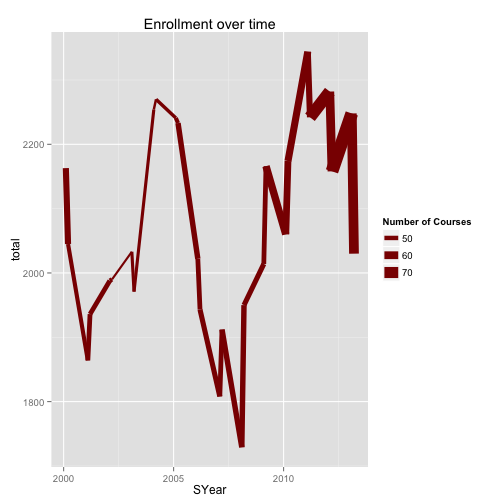


Testing the difference of the locations on enrollment, using dummy variables for both year and location, shows that there are significant differences in enrollment patterns for the Main and Online campuses. All other campuses do not show significant variation.

## Other Questions that have been asked

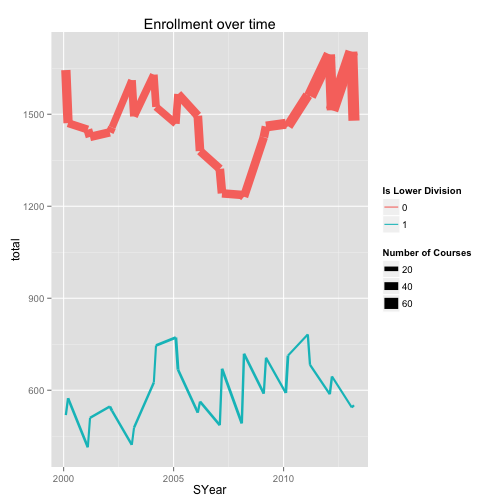
#### If we offer more classes/sections are we competing for the same students?

We can take a look at a modified version of the above graph showing enrollment over time and change the thickness of the line to show how many course are being offered. There doesn't appear to be much correlation between the number of courses offered and enrollment



A basic regression with total number of students as the D.V. and the number of courses as the I.V. doesn't have any significant effects, not does it fit the data very well with an r2 value of 0.1156. Controlling for year doesn't seem to help either.

However, if we analyze the number of students in a course based on the number of sections that are offered, and control for the course being lower division (because SOC 1010 throws everything off) then we get a positive correlation and an r2 of 0.5477



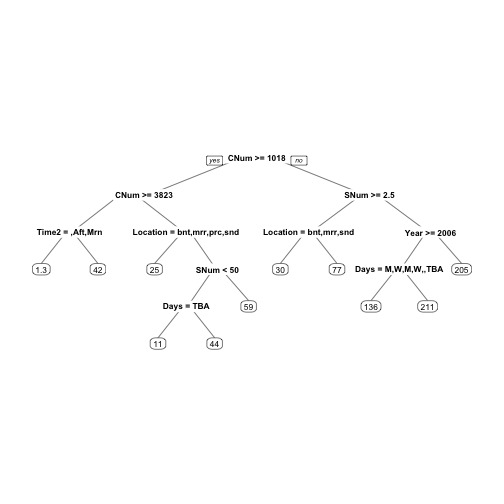
The results from a linear regression model suggest that increasing the number of sections of a course correlates with higher enrollment even when accounting for the difference between lower and upper division courses.

### What Does Enrollment look like without branches

There have been a total of 44071 students in the taking courses on the main campus since 2000. 8153 were online and the remaining 5846 were from the branches. If there were no courses offered at the branches and those students did not come up to campus or attend online there would be a decrease of 10.0672%. However if we only look at the data since 2010 then the decrease is only 3.8032%, or 667 students.

### What Aspects of Courses or Schedules Generates Higher Enrollment?

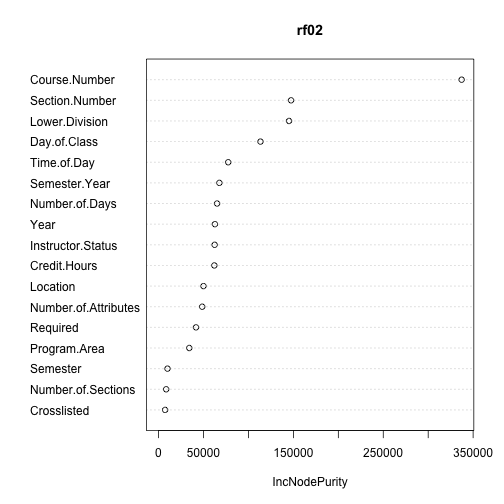
A number of questions asked deal directly with what factors generate higher enrollment. The best technique for this type of question (where there is no direct theory to test) is a Classification and Regression Tree (CART) data mining algorithm. This algorithm takes the available data and partitions it into similar clusters. If we throw all of the data into a cart model we get the following picture.



The decision tree aspect of CART models allows for easy interpretation, each branch of the tree is a variable that has an impact on enrollment. The circles at the bottom of the tree are the average enrollments for courses following the path of the branches. So the highest enrolling courses are those with a course number less than 1018, with a section number less than 2.5 since 2006 that is held on Tuesdays and Thursdays with an average of 206 students.

While this is a nice representation of the data we currently have it may not help much in predicting future course enrollments due to the over fitting of the model. A better choice is the random forest model. The random forest model takes random samples of the data set,runs the cart model on that sample and uses the results to predict the values in the rest of the data set. This process is repeated over and over again to eliminate the influence of any outliers and results in a robust predictive model.

This model had a mean squared error of 310.3706 and explains 3.1 × 104% of the variance after 70 iterations. The plot below indicates the variables that had the greatest importance. The five most important variables are course number, being lower division, the section number, days class is held, and the number of credit hours in a course.



From this analysis the number of attributes a course has, or the specialization area don't have much influence in the course enrollment.

We can use this model to predict the enrollment for next semester by feeding the model the Spring 2013 course schedule. The model predicted 2130 students to be enrolled, while 2080 actually enrolled, a difference of 50 or 2.4% over prediction. The actual amount of over prediction was a bit higher because the model could not predict the enrollment for several courses due to a lack of data.

Here is the course by course breakdown of Spring 2014 courses:

### Spring 2014 Schedule

|  |  |  |  |
| --- | --- | --- | --- |
| Course.Number | Section.Number | Enroll | Predict |
| 1010 | 1 | 87 | 141.01 |
| 1010 | 2 | 144 | 179.56 |
| 1010 | 70 | 14 | 25.16 |
| 1010 | 90 | 58 | 105.60 |
| 1015 | 90 | 78 | 88.61 |
| 1020 | 20 | 5 | 16.85 |
| 3041 | 1 | 91 | 68.89 |
| 3042 | 1 | 58 | 35.78 |
| 3061 | 90 | 49 | 55.88 |
| 3111 | 90 | 55 | 61.09 |
| 3111 | 1 | 42 | 37.15 |
| 3112 | 90 | 72 | 76.02 |
| 3112 | 1 | 45 | 63.85 |
| 3112 | 10 | 85 | 44.36 |
| 3140 | 1 | 22 | 38.61 |
| 3140 | 90 | 78 | 68.92 |
| 3337 | 1 | 38 | 42.64 |
| 3365 | 1 | 10 | 35.68 |
| 3380 | 90 | 64 | 50.89 |
| 3393 | 1 | 5 | 13.33 |
| 3435 | 60 | 3 | 11.85 |
| 3440 | 90 | 80 | 67.62 |
| 3446 | 90 | 53 | 59.61 |
| 3480 | 2 | 3 | 29.05 |
| 3480 | 1 | 32 | 33.38 |
| 3486 | 1 | 6 | NA |
| 3510 | 1 | 46 | 34.27 |
| 3560 | 1 | 105 | 53.64 |
| 3561 | 90 | 91 | 80.60 |
| 3562 | 90 | 35 | 67.24 |
| 3565 | 1 | 73 | 59.37 |
| 3569 | 90 | 81 | 62.15 |
| 3570 | 10 | 37 | NA |
| 3593 | 1 | 19 | 17.18 |
| 3638 | 70 | 3 | 17.53 |
| 3650 | 90 | 124 | 75.56 |
| 3671 | 1 | 14 | 29.03 |
| 3671 | 2 | 3 | 19.13 |
| 3673 | 90 | 107 | 88.36 |
| 3741 | 1 | 8 | 36.92 |
| 3769 | 1 | 13 | 37.36 |
| 3950 | 1 | 0 | NA |
| 3950 | 3 | 1 | NA |
| 3950 | 5 | 0 | NA |
| 3950 | 6 | 0 | NA |
| 3950 | 7 | 0 | NA |
| 3951 | 3 | 1 | NA |
| 3951 | 5 | 0 | NA |
| 3951 | 1 | 0 | NA |
| 3952 | 1 | 0 | NA |
| 3952 | 6 | 13 | NA |
| 4085 | 1 | 10 | NA |
| 4438 | 1 | 3 | NA |
| 4693 | 7 | 1 | NA |
| 4693 | 8 | 2 | NA |
| 4693 | 1 | 0 | NA |
| 4693 | 2 | 1 | NA |
| 4693 | 3 | 0 | NA |
| 4693 | 6 | 1 | NA |
| 4693 | 4 | 0 | NA |
| 4877 | 1 | 0 | NA |
| 4877 | 2 | 1 | NA |
| 4877 | 3 | 0 | NA |
| 4877 | 4 | 0 | NA |
| 4877 | 5 | 1 | NA |
| 4998 | 1 | 1 | NA |
| 4999 | 2 | 0 | NA |
| 4999 | 1 | 0 | NA |
| 4999 | 3 | 1 | NA |
| 4999 | 4 | 0 | NA |
| 5440 | 90 | 7 | NA |

## Conclusion

It appears that branch campuses have declined in enrollment since the advent of online courses. There also seems to be a decline in enrollment at the main campus as well. Online courses are the only campus that has seen an increase in enrollment over the past few years. In linear regression models the variation over time did not apper to be statistically significant.

From the CART models it appers that after accounting for the high enrollment in Soc 1010, the most important factors for high enrollment are location, either at the main campus or online.