# A Case Study: Exploring Student Academic Performance Data for Actionable Knowledge

Omar ElTayeby, Carlos E. Seminario, Wenwen Dou, Mary Lou Maher and Evan Murphy, *UNC Charlotte*

**Abstract**— For decades, university and college administrators have been concerned with student academic performance regarding time-to-degree and time-to-attrition as well as determining the reasons why students take too long to graduate or leave school before receiving their degree. Although this problem has been researched extensively, the concerns remain to this day. With the advent of big data, data science, and sophisticated visualization approaches, there are more tools at our disposal to analyze and explore student data in the hopes of better identifying or predicting which students are at-risk. However, determining which features to mine remain a challenge for this type of analysis. In this case study, we analyzed 10 years of student data in the College of Computing and Informatics (CCI) of a large state university. Using EventFlow, a tool that analyzes event sequences, we developed a novel knowledge generation approach that can be used to test hypotheses about reasons for student attrition.

**Index Terms**—Learning analytics, student data, hypothesis generation

✦

## 1 INTRODUCTION

Academic leaderships have spent much effort on increasing retention and reducing attrition rates at universities. Our goal is to explore how interactive visual analytics can facilitate hypothesis generation by subject matter experts, who aim to understand the reasons behind attrition and delayed graduation. We use data collected over 10 years from UNC Charlotte (UNCC) database which holds records about student?s academic performance. In prior research, researchers have focused on developing tools for providing actionable knowledge, such as Course Signals [1], to solve problems which cause high rates of attrition. We develop an approach to address decision makers' challenges from the data science perspective. Our approach will enhance the knowledge discovery for decision makers by placing them in the center of the sense-making process [2]. We employed EventFlow [3], a visual analytics system for analyzing sequence data, to explore the student's progress over time and find the important features for building predictive models. We found that EventFlow is very helpful in generating hypotheses and cleaning incomplete records.

Eventflow is an interactive tool for visualizing sequence data. It consists of three panels: control, overview and timeline. The tool provides encodings for two types of events: point and interval. The overview panel aggregates similar events to align them regardless of the exact times of occurrences. The aggregation is helpful for data analysts to see the mean and median times elapsed between consecutive events. The panel also shows the general patterns, which indicates the significance of repeated sequences. Our demonstration goal is to provide an opportunity for the workshop audience to explore our anonymized student dataset using EventFlow, through which we will collect feedback about the different scenarios of exploration styles and hypothesis generation. We will present visualizations using EventFlow to demonstrate 3 hypotheses which were observed from our collaboration with faculty leaders.

## 2 DATA EXTRACT AND CLEANING

We conducted a time-to-degree and time-to-attrition analysis of CCI undergraduate students with the intent of determining not only the

---

- *Omar ElTayeby is with UNC Charlotte. E-mail: oeltayeb@uncc.edu.*
- *Carlos E. Seminario is with UNC Charlotte. E-mail: cseminar@uncc.edu.*
- *Wenwen Dou is with UNC Charlotte. E-mail: wdou1@uncc.edu.*
- *Mary Lou Maher is with UNC Charlotte. E-mail: mmaher9@uncc.edu.*
- *Evan Murphy is with UNC Charlotte. E-mail: emurph26@uncc.edu.*

---

amount of time required to graduate from, or leave, CCI but also to determine whether there were any differences in these durations based on student gender. Ten years of student data (2004-2014) were extracted and filtered from UNC Charlotte's Operational Data Store (ODS). The working dataset for this analysis consists of 4814 students, of which 3985 were male and 829 were female.

In our preliminary data filtering and cleaning, we excluded the students who were denied for CCI admission or never enrolled, and included the ones who majored in CCI for at least one semester during their undergraduates studies. During our analysis using EventFlow we identified the students who had incomplete records, when we observed missing activities inside the sequences. The presence of those incomplete student records could skew the results. EventFlow helped us to specify the filtering conditions needed to exclude those students, and then iteratively verify them when looking at individual students' sequences in the timeline panel.

## 3 PRELIMINARY RESULTS

In order to derive statistics on the time-to-degree and time-to-attrition, in our primary visualization, we represent the events of admission, graduation, changing majors, and attrition as point events. The courses taken per semester are represented as interval events. Using the overview panel shown in Fig. 1, we found that:

- The median time-to-degree

  - Was about 4-5 academic years for new freshmen CCI admissions. These same durations were observed when broken down by gender
  - Was about 3-4 academic years for all of students admitted to CCI (including transfer and special students); 35 months for male students and 40 months for female students

- The median time-to-attrition

  - Was 9 months to leave UNCC and 1 year to leave CCI (changed to a non-CCI major) for new freshmen CCI admissions. These same durations were observed when broken down by gender
  - Was 9 months to leave UNCC and 1 year to change from CCI to another program in UNCC for all of the students admitted to CCI (new freshmen, transfers, special students, etc.)

We observe that the reduction in time-to-graduate for the CCI students admitted at large is likely due to a large number of transfer
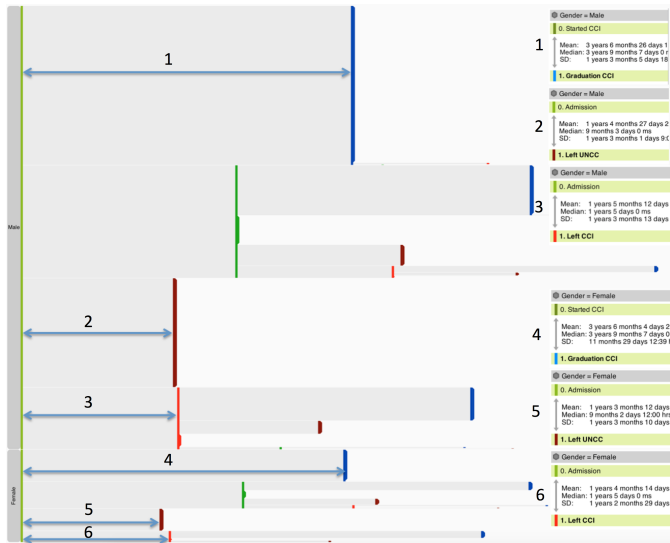
Fig. 1. The overview panel showing the time to graduation, leave CCI and UNCC for male and female students.

students that take fewer years to graduate than new freshmen. In conclusion, the statistics on both genders are insignificantly different in terms of median time to leave CCI or UNCC.

When we interviewed the CCI's academic leaders to explore the dataset using EventFlow, they identified 3 major hypotheses which warrants further validation. They liked the capabilities of EventFlow that permit interactive analysis of student data. During the interview, we first added the main events which concern their general queries, and then the events that address their specific hypotheses. Such visualization setup enables them to observe and compare the general patterns of sequences. In order to avoid clutter in the overview panel we started with an empty overview then added events that only relate to the query. On the other hand, we display all events in the timeline panel to view more details for selected students from the overview panel. Secondly, we were able to show them statistics about the distribution of the time elapsed between consecutive events. Additionally, when focusing on particular time distributions, EventFlow gave us the ability to accommodate their queries that require selecting certain students within their specified time window. Lastly, we used the timeline panel to verify the aggregate patterns by taking a closer look at individual students and details that are not presented in the overview.

## 4 HYPOTHESES GENERATED

Hypothesis 1: "The reasons for female attrition are different than male attrition." As mentioned earlier in preliminary results, there are differences in the statistics between male and female for time-to-attrition. This difference lead to a question about the reasons for female students leaving CCI, and whether they are different from the attrition reasons for male students. One of the team members was interested in querying the GPA of the female students who leave CCI, in order to know the percentage of high achievers (with GPA greater than 3) who decided to leave. By incorporating the GPA information, we divide the cumulative GPA into 3 categories: less than 2, between 2 and 3, and greater than 3. When splitting the CCI semester events (represented as intervals) by GPA, we were able to find the percentages of freshmen students who had GPA greater than 3 and changed their major, which were 31.6% for male and 37.5% for female students. While the freshmen students who left UNCC with a GPA greater than 3 were 14.5% among the male and 13.8% among the female students, which indicate insignificant differences. We showed the progression of GPA to better understand which courses made the students fail or lose interest. The CCI and non-CCI courses are visualized with the same color in two different shades and each interval event indicates the number of courses and the grades for each course. The combination of

the overview panel that enables the query of interesting patterns and the timeline panel that shows more details about individual students see useful for exploring patterns of different student groups.

Hypothesis 2: "Some students apply for CCI majors as a stepping stone to get accepted in UNCC then they transfer to another major." One of the faculty members asked to find out whether there is a significant number of students who applied for CCI majors then change the major without taking any courses, just to get into UNCC. We included the CCI courses as interval events in the overview and timeline panels to search for the portion of students who did not take any CCI courses before changing their major. It was easy to spot the students with the specified criteria and we found that only 8 students in the collected 10 years of student records used CCI as a stepping stone towards being accepted. The faculty member did not consider students with the same criteria who were already enrolled in a non-CCI major before getting into CCI; we excluded them by adding the acceptance in non-CCI colleges as point events in the overview panel.

Hypothesis 3: "Staying the same level of courses is a good predictor for detecting students at risk." One of the researchers on the team is building a predictive model to detect the students at risk during the Sophomore year. Since the feature engineering is very important in building the model, she hypothesized if we can rely on the student's course progress as a good predictor. In order to validate this hypothesis, we represented courses of the same level for each semester in one event interval, where each course level is represented with a different color. Also the event intervals include details about the course names and grades, when hovering on them in the timeline panel. The sorting of the events in the control panel is helpful to visually indicate the progress of the course levels along time; we sorted the lower and higher level course to be at the bottom and top respectively. When we included the different course levels in the overview panel, it became cluttered with many combinations of patterns, so we used the merge capability to clear out gaps and overlaps within one year intervals. Out of the freshmen students who leave CCI to another college, one third of them take a one thousand level course then leave without taking course from other levels. This significant portion of students is an indication that the course level change could help in improving the prediction model for detecting students at risk.

## 5 CONCLUSION

In this case study we have shown that the use of the time-series visualization tool, EventFlow, can enhance data analysis efforts to better identify or predict the students who are at-risk of attrition. We were able to quickly generate and test hypotheses with subject matter experts and leaders.

## REFERENCES

[1] Kimberly E. Arnold and Matthew D. Pistilli. Course signals at Purdue: using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), Simon Buckingham Shum, Dragan Gasevic, and Rebecca Ferguson (Eds.). ACM, 267-270, 2012.

[2] Monroe, M., Lan, R., Morales del Olmo, J., Shneiderman, B., Plaisant, C., Millstein, J. *Proc. Of ACM Conference on Human-Computer Interaction* (CHI 2013), 2012.

[3] Monroe, M., Lan, R., Lee, H., Plaisant, C., & Shneiderman, B. Temporal event sequence simplification. *IEEE transactions on visualization and computer graphics*. 2227-2236, 2013.