*Do staffing levels and salary outlays of different types of employees in higher education have an impact on completions?*

The purpose of this research is to identify any relationships between how much college and university employees are paid, how many are employed, and how many completions there are. Completions are defined as the number of students receiving awards/degrees. This dataset is awards/degrees conferred between July 1, 2018 and June 30, 2019 to all recipients, across all race/ethnicities and both genders.

Data is also included for Instructional Staff, Non-Instructional Staff, and a subsection of Non-Instructional Staff classified as Education Services. These would include people working in libraries, museums, and educational service positions. Details are provided in the list of variables from the data dictionary.

Awards are separated into Associate’s, Bachelor’s, Master’s, and Doctoral degrees.

The first result of my EDA was that I needed to apply a transformation to my variables. This would normalize the distribution of my data and allow me to test it. Some additional exploration led me to experimenting with calculated salary averages by dividing the total salary outlay by the number of employees. This data did not need to be normalized, the distribution was already normal when plotted. It makes sense that this would be a smoother plot, as averaging the values out across the set would smooth out the extremes. This average salary data did not end up being useful, with an R2 of .19.

I think I could have added more in the analysis like other types of non-instructional staff. I chose educational support staff because that group had more likelihood of directly impacting students than the other subgroups based on the definitions in the dataset. I think looking specifically at Management, Sales, and Office and Administrative Support groups might provide a more thorough look at places universities might spend a large amount of capital that would not be directly tied to student success.

The variable HDEGOFR1 listing the highest degree offered by a given institution might have been good to use when breaking down salary ranges, as a school offering Doctoral degrees probably has a larger staff budget than a small community college with Associate’s degrees. There was also an academic rank variable for instructional staff that broke them down into seven categories. I chose not to include these because the categories are somewhat arbitrary and schools choose where their ranks fit into the scale. The available data is also broken down by male and female staff, so a later study could be done looking at equity and salary.

I assumed when starting this project that non-instructional staff pay and counts would outstrip that of instructional staff, and it seems to be generally correct. But I did not expect educational support staff to fall below instructors. There were some NA values for that data, so it is possible schools classify their support staff differently. I also assumed that using the mean salaries I calculated would be more useful in a model but that was not the case. The overall salary outlay would cover a broader spectrum than the average in trying to gauge the economic capital a school is investing on students.

It was challenging to create some of my plots, having them laid out properly with or without certain chart features. I also had difficulty working with my data because it required log transformation. Many of my hypothesis tests only returned a 0.0 but I do not know if that was due to the test not accounting for the log transformation or a true result. The plots seem to show a relationship, but I do not know if I am just forcing the data into what I want to see at the cost of validity.