Data Source: <https://www.deltacostproject.org/delta-cost-data>

Research Question: How do the ways in which public 4-year institutions spend their money affect graduation rates?

Data Exploration: We reviewed the data dictionary (provided with the data) and identified key variables we would be interested in examining/including in our models. For those variables, we examined the yearly missing rates to identify which ones would be unusable and which ones could be workable with some minor imputation.

Literature Review: We reviewed two research papers/projects that used this data to examine the impact of Outcomes (or Performance) Based Funding on bachelor’s degree attainment in public 4-year institutions to get ideas on how to define our analytical sample, which key variables we should include in our models, and how those who work in higher-education research standardize key variables to control for institution size. The papers are below:

* <http://journals.sagepub.com/doi/pdf/10.3102/0162373714560224>
* <https://aefpweb.org/sites/default/files/webform/42/Playing%20the%20Long%20Game%20-%20Slaughter%20et%20al..pdf>

Data Preparation: We first found and cleaned data on state and year specific GDP per capita so they could be merged with our final dataset, as the two studies above included them in their models. Because graduation rates are only available starting in academic year 2003, we started by keeping years past then in the data, which left us with 553 institutions. We then thinned our sample out in the following way:

* Kept schools in the 50 states plus DC (544 institutions kept)
* Dropped schools that didn’t exist in the entirety of the time range, academic years 2003-2015 (505 institutions kept)
* Dropped schools that were missing bachelor’s degrees awarded or graduation rate in any of the years in the data (457 institutions kept)
* Dropped schools that gave out 0 bachelor’s degrees in a year or awarded fewer bachelor’s degrees than associate degrees in a year (445 schools kept)

We then standardized monetary variables to 2015 dollar values using CPI scalars and standardized variables that would be sensitive to institution size by dividing them by the number of fall full-time equivalent (FTE) students.

The final step for preparation is to determine how to impute missing values for variables with workable missing rates.