**Data Summary**

Longitudinal data for every higher education institute in USA with information on finances, enrolment and student outcomes, for each year from 1987 to 1999, and 2000 to 2015 in two data files.

2000 -2015 data has 672 variables and 43171 observations. 234 variables are “flag” variables denoting if data was imputed – so only 438 “real variables”.

**General Ideas:**

General themes of interest – are different methods of funding and expenditure associated with different outcomes for students. Is there an optimal way to expend/fund for given goals – e.g. research or general education.

* Possible of interest variables –
  + How colleges are funded – finance variables.
  + How colleges spend their money – expenditure variables.
* Possible outcomes –
  + Graduation Rates.
  + Total completions per 100 FTE students
  + Number of master, bachelors and doctoral degrees and proportions.

**Summary of Available Covariates**

The below table summarises the covariates, with each one have typically have many combinations of associated variables:

|  |  |
| --- | --- |
| **Covariate** | **Description** |
| Institutional characteristics | General information about institution |
| Completions |  |
| Enrolments |  |
| Fall Staff |  |
| Finance - Expenditures | Split by salaries, different sectors, includes depreciation |
| Finance - Revenues | Split in many different ways. |
| Finance - Scholarships |  |
| Graduation Rates |  |

**Initial Look at Data Ideas**

* Restrict our analysis to 4 year public institutions, since otherwise there will be many heterogeneous funding sources.
* Standard regression for graduation rates
* Count model – Poisson regression for number of bachelor degrees awarded.
* Look at number of doctoral, bachelor, and masters degrees awarded – only on subset of doctoral research universities.
* We have a variety of indices : CPI, HEPI, HECA, the former being general prices the latter applying to higher education in particular. Should rerun analysis with each index to se to see if it is inflation index dependent – suspect could provide different results.
* Various different tuition/aid/grants etc – likely a lot of multicollinearity between these, could include all, some or just gross tutition. Have these in absolute terms.
* Also have tuition reliance and government reliance i.e. share of operating revenue derived from tuition or government.

**Possible of interests Variables**

* Percentage of undergrads receiving financial aid. - **OUTCOME**
* Percentage of undergrads seeking student loans – is this related to tuition size or other variables. - **OUTCOME**
* Expenditure – how is this related to tuition, percentage of government aid – for example are institutions with high /low levels of government aid incentivised to cut costs/spend efficiently, maybe we could make some sort of comparison to private universities.
* Salaries as a proportion of instruction expenditure – is there a relation? – also salaries as a proportion of other expenditures – which are favourable?
* Are different categories of expenditure related to numbers of masters, doctorate, bachelor degress awarded i.e. does increased spending on research increase number of doctorate degrees?
* Does student service spending help – or is it just white elephant spending?
* Depreciation – may indicate colleges with big capital spending – does this help?