This file summarises the processing that has been done to the data to create our sample.

1. Removed non 4 year institutions ( sector\_revised variable)
2. Checked groupings – used groupid as unique identifier – not much difference to unitid
3. Removed imputation variables
4. Removed location variables not needed.
5. Dropped sector levels – not needed
6. Checked missingness on variables of interest – concluded use data post 2003
7. Scaled dollar values by CPI to 2015 dollars.
8. Concluded we should use fte\_count – fall enrollment not 12 month enrollment, since less missingness
9. Made final sample with required variables
10. Removed institutions not primarily awarded bachelor degrees, included only years after they awarded at least 10 bachelor degrees – somewhat arbitrary
11. Dropped universities that have missing graduation rates – somewhat worrying removes 30 institutions.
12. Final sample called sample\_2003