Our files from CDE go back to 2012-13 to the present. Go through and pull together historical enrollment by grade level, institution, and state so that I have a nice panel data set to validate our internal series against.

I am not particularly interested in race/gender but that would be a plus

<https://www.cde.ca.gov/ds/ad/filesenr.asp> ( This goes back further but it is a census day)

<https://www.cde.ca.gov/ds/ad/filesenrcum.asp> ( I know the time series limitation here to start in 2016-17)

I would need you folks to do the same exercise for graduation using the the two times

<https://www.cde.ca.gov/ds/ad/filesoygrads.asp>

<https://www.cde.ca.gov/ds/ad/filesgrad.asp>

I also need the same data for IPEDS with a focus on Institution Level Enrollment and By Major/Degree Awards for UC, CSU, and CCC institutions.

The time frame is 2011-2023.

<https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx?gotoReportId=7&fromIpeds=true&sid=14b024c2-b240-492a-a4d5-67b1cdd3e35e&rtid=1>

For all of these, I would like this code to be replicable and modifiable. Please write the scripts into an R project or even have it posted on a github where I could pull all of this down.

The use cases I wanted to capture and articulate in the white paper are to enumerate our choices in resolving race:

1. Use some primacy rules
2. When things conflict, pool that information (e.g.  White/Hispanic -> Two or more races)
3. Contextual information (treat information provided to  a social services worker as different as received from CDE or other agencies)