Capstone Project 1 Data Story

Purpose

The purpose of this document is to provide information necessary to appropriately use school and district level data files from the state of Massachusetts for 2011, 2012, 2013, 2014, 2015 years. The data contains regulatory adjusted-cohort graduation rates (ACGR) from EDFacts, and is crucial to take into consideration prior to conducting any analyses on the data.

EDFacts Background

EDFacts is a Department of Education (ED) initiative to govern, acquire, validate, and use high-quality elementary and secondary performance data in education planning, policymaking, and management decision making to improve outcomes for students. EDFacts centralizes data provided by the state education agencies (SEAs) at the state, local education agency (LEA), and school levels (SCH). EDFacts also provides the Department with the ability to easily analyze and report the data. Since its inception in 2004, this initiative has reduced reporting burden for SEAs and local data producers, and has streamlined elementary and secondary data collection, analysis, and reporting functions at the federal, state, and local levelsSelected Criteria

Table 1. ED*Facts* Four-Year ACGR File Specifications and Data Groups

File Specification	Data Group	Data Group Name	Data Group Definition
FS150	DG695	Regulatory four -year adjusted-cohort graduation rate table	The regulatory four-year adjusted-cohort graduation rate is the number of students who graduate in four years with a regular high school diploma divided by the number of students who formed the cohort for that graduating class. The four-year adjusted cohort rate also includes students who graduate in less than four years.

The graduation rates are reported in the following subgroups, as required by law:

- Major Racial and Ethnic Groups
- Disability Status
- LEP Status
- Economically Disadvantaged Status

Refer https://www2.ed.gov/about/inits/ed/edfacts/index.html to access the file specifications.

Education Levels Reported

States submit data at three education levels: SEA, LEA (includes school districts), and SCH. Each LEA is assigned a 7-digit ID by the National Center for Education Statistics (NCES). The first two digits represent the state and the last 5 digits are unique within that state for the LEA. Each school is also assigned a unique ID by NCES. The school IDs are 12 digits. The first 7 digits represent the LEA that the school belongs to and the remaining 5 digits are unique to that school within the LEA. However, while the remaining 5 digits may not be unique within the state, the entire 12-digit school ID is unique within the state and the nation.

The following formula provides an example of how the four-year adjusted cohort graduation rate would be calculated for the cohort entering 9th grade for the first time in the 2011-12 school year and graduating by the end of the 2014-15 school year:

Number of cohort members who earned a regular high school diploma by the end of the 2014-15 school year

---

Number of first-time 9th graders in fall 2011 (starting cohort) plus students who transferred in, minus students who transferred out, emigrated, or died during school years 2011-12, 2012-13, 2013-14, and 2014-15

Table 2. Subgroup Abbreviations

Abbreviation	Meaning
ALL	All students in the school
	Major racial and ethnic groups representing:
MAM	American Indian/Alaska Native students
MAS	Asian/Pacific Islander students
MHI	Hispanic students
MBL	Black students
MWH	White students
MTR	Two or More Races
CWD	Children with disabilities (IDEA)
ECD	Economically disadvantaged students
LEP	Limited English proficient students

[METRIC]: All data are aggregated by subgroup. For each subgroup within the file there are two metrics presented in the ACGR files:

Table 3. Abbreviations for metrics in ACGR file

Abbreviation	Meaning
COHORT	The total number of students within the adjusted-cohort (the sum of both graduate and non-graduate students)

Data Analysis

The graduation data for years 2011-2015 are read in as CSV file and the following were performed -

1. The school code was converted from float to int while reading in the data.
2. Filter was applied to read in one state.
3. SUBGROUP LEP_RATE - Filtered out the columns that are not in scope for analysis-that is, include only RATE columns
4. SUBGROUP MTR_RATE - Analysis was done for each value in MTR_RATE, to count the number of occurrences excluding NULLS
5. SUBGROUP CWD_RATE - Filtered out to include only RATE columns; further filter out non null CWD_RATE values; customized output sorted by mean values
6. SUBGROUP ECD_RATE - Filtered out to include only RATE columns; further filter out non null ECD_RATE values; setting and removing an index; resetting index can help remove hierarchical indexes while preserving the table in its basic structure; sort the resulting column by index

References

https://nces.ed.gov/pubs2018/2018117.pdf
https://nces.ed.gov/ccd/drp7yrag.asp
https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=REL2018277
https://nces.ed.gov/ccd/elsi/tableGenerator.aspx?savedTableID=72529#
https://www.edweek.org/ew/articles/2017/12/07/whats-behind-the-record-rises-in-us.html
https://www.edweek.org/ew/section/multimedia/data-us-graduation-rates-by-state-and.html
https://www.edweek.org/ew/articles/2017/12/07/whats-behind-the-record-rises-in-us.html
https://www.edweek.org/ew/section/multimedia/us-education-in-2017-in-10-charts.html
https://nces.ed.gov/datalab/powerstats/pdf/els2002sophomores_subject.pdf
https://nces.ed.gov/Datalab/QuickStats/Output
https://www.kaggle.com/askhanna/student-dataset-with-graduation-details#student-data.csv
http://profiles.doe.mass.edu/statereport/gradrates.aspx
https://www.youtube.com/watch?v=kiwSwuic0TA&list=PLTuqmiQ9ssqsCEVNSi2Zm2FPHqZoHLAuu
http://profiles.doe.mass.edu/grad/grad_report.aspx?orgcode=00000000&orgtypecode=0&&fycode=2017
http://profiles.doe.mass.edu/statereport/gradrates.aspx