Data and their sources

Data sets, aggregated by school district:
[1] ACGR School Year 2012-2013 Data
http://www2.ed.gov/about/inits/ed/edfacts/data-files/acgr-lea-sy2012-13.csv

[2] NCES 2013 Data for Income and Poverty
http://nces.ed.gov/programs/edge/tables.aspx?ds=acsProfile&y=2013

[3] ElSi tableGenerator data (I created two sets on separate occasions and merged the results in R):
http://nces.ed.gov/ccd/elsi/tableGenerator.aspx?savedTableID=16594
http://nces.ed.gov/ccd/elsi/tableGenerator.aspx?savedTableID=36934

Variables and descriptions:

Since it is not uncommon for students to enter and exit school districts each academic year, **4-year adjusted cohort graduation rates (ACGR)** will be the version of high school graduation rates that is studied in this investigation since it accounts for student mobility, including transfers, emigration, and death during the 4-year academic period. Because of the way that it is calculated, ACGR is considered to be an accurate estimate of 4-year graduation rates. The **ACGR** data are available from the U.S. Department of Education, EDFacts Data Files website. This data set corresponds to School Year 2012-2013, is indexed with a FIPS identifier for each school district in all 50 U.S. States, including the District of Columbia, and contains 11,957 records. It contains graduation rates for all students, and the data is disaggregated by ethnicity.

The primary predictor that will be included in this investigation is **Instructional Expenditure per Student**. Instructional expenditures include costs for teacher salaries, benefits, and instructional services.

Median Household Income data will be collected from the Education Demographic and Geographic Estimates (EDGE) website, using the available Table Viewer tool. The data are derived from the U.S. Census Bureau American Community Survey and are called "ACS Profile Tables 2009-13". These 13,411 records are indexed by school district using FIPS codes. Income likely plays a role in 4-year public high school graduation rates since it impacts the resources available to a student, such as shelter, food, transportation, materials for school, etc. On the other hand, a student's humble (prosperous) economic background could potentially serve as strong (weak) motivation to earn a high school diploma.

Another measure that could be included for each district, and which comes from the aforementioned EDGE website is **"percentage of families and people whose income in the past 12 months is below the poverty level"** since this could be another general measure of a school community's economic power.

Other data that will be included in this investigation are **Urban centric locale indicator** (several levels), **Pupil/Teacher Ratio, Total Free Lunch Eligible and Reduced-price lunch**

eligible students (as an alternate measure of poverty), Limited English Proficient / English Language Learners & Total Enrollment, Total Number of District Staff, and Student Ethnicity. These data can be found by using the Elementary / Secondary Information System (ElSi) tableGenerator tool on the National Center for Education Statistics website. These 13,619 data records are also indexed by FIPS code.

Variables that are currently not included but would be interesting:

If it is available somewhere, I will also include data for **"average daily attendance"**, since attendance is likely to have an impact on 4-Year Public High School Graduation Rates. Additionally, if I can find data on **average of highest education level achieved by parents**, I will include that as well.