Schema Design (College Scorecard Dataset)

Vinay Maruri, Shane Hauck, Xiaowen Wu

```
In [ ]:  CREATE TABLE InstitutionInformation (
             unitid INTEGER PRIMARY KEY,
             instnm TEXT,
             location TEXT,
             address TEXT,
             region INTEGER,
             control INTEGER,
             ccbasic INTEGER,
             censusids INTEGER,
             latitude INTEGER,
             longitude INTEGER,
             accredagency TEXT,
             preddeg INTEGER,
             highdeg INTEGER,
             avgfacsal INTEGER
         )

         CREATE TABLE StudentBody (
             unitid INTEGER PRIMARY KEY,
             sat_avg INTEGER,
             adm_rate FLOAT,
             pptug_ef FLOAT,
             ugds_white FLOAT,
             ugds_black FLOAT,
             ugds_hisp FLOAT,
             ugds_asian FLOAT,
             ugds_NRA FLOAT,
             ug INTEGER,
             inexpfte INTEGER,
             C150_4 INTEGER,
             C150_L4 INTEGER,
             tuitfte INTEGER,
             tutionfee_in INTEGER,
             tutionfee_out INTEGER,
             tutionfee_prog INTEGER,
             FOREIGN KEY (unitid) REFERENCES InstitutionInformation(unitid)
         )

         CREATE TABLE Debt (
             unitid INTEGER PRIMARY KEY,
             grad_debt_mdn INTEGER,
             wdraw_debt_mdn INTEGER,
             lo_inc_debt_mdn INTEGER,
             md_inc_debt_mdn INTEGER,
             hi_inc_debt_mdn INTEGER,
             dep_debt_mdn INTEGER,
             ind_debt_mdn INTEGER,
             pell_debt_mdn INTEGER,
             nopell_debt_mdn INTEGER,
             female_debt_mdn INTEGER,
             male_debt_mdn INTEGER,
             firstgen_debt_mdn INTEGER,
             notfirstgen_debt_mdn INTEGER,
```

```
    CDR2 INTEGER,
    CDR3 INTEGER,
    FOREIGN KEY (unitid) REFERENCES InstitutionInformation(unitid)
)

CREATE TABLE StudentOutcomes (
    unitid INTEGER PRIMARY KEY,
    md_earn_wne_6 INTEGER,
    pct25_earn_wne_p6 INTEGER,
    pct75_earn_wne_p6 INTEGER
    count_wne_inc1_p6 INTEGER,
    count_wne_inc2_p6 INTEGER,
    count_wne_inc3_p6 INTEGER,
    FOREIGN KEY (unitid) REFERENCES InstitutionInformation(unitid)
)
```

Comments:

The basic entities in our schema are InstitutionInformation, StudentBody, Debt, and StudentOutcomes. We chose these four entities because we felt like they best described the information contained within the college scorecard database at the student and institution level. InstitutionInformation summarizes key information about the colleges/universities themselves in terms of identifying them, and telling us where they are, what degrees they offer, and other information about the university itself. StudentBody tells us overall information about the student body at each institution such as demographics, what their scores are, how many part time studnts there are, how much is spent on each student, and information about how much tuition students pay and what their degree completion rates are. Debt tells us summary statistics for how indebted students at each institution are (broken out by demographics) and what student default rates are. StudentOutcomes tells us about how students at each institution do after graduating or leaving school in terms of their salaries and employment outcomes.

There is no redundant information in our database, as no information is repeated between tables except for the foreign keys linking them all together at the institution level. All information in each table is distinct from each other and does not overlap outside of the institution level. We made sure of this, and this is reflected in our schema design.