# Data Collection Report

Jianing Yang

## Location of Database Table

Database: schools4_database
Schema & Table: jianing_yang_acs_schema.jianing_yang_acs_table

## Code

Repository: https://github.com/jedyang97/acs_data_download
See README.md for instructions to install and run the code. Getting data, transforming data and loading data are all automated by scripts and user can just **run three shell scripts** to complete all of these actions. The code is modularized and generalizable so that it can download data at different geography level and add additional variables (see example in get_data.sh for how to do this).

## Data

**Where to get data?**
I chose to use https://api.census.gov/ as the source of data, because this is the official government's website, which constitutes an authoritative, reliable and up-to-date data source.
I used the API here: https://api.census.gov/data/2018/acs/acs5/, by following the documentation here:
https://api.census.gov/data/2018/acs/acs5/examples.html?eml=gd&utm_medium=email&utm_source=govdelivery.

**What variables to use?**
I selected 12 variables based on 4 categories: race, income, gender and education. Variables in these categories are most pertinent to our high school graduation intervention project because I think there could be a high correlation between this family background information and the success of a student in high school.

| Race | Total | B02001_001E |
|------|-------|-------------|
| | White alone | B02001_002E |
| | Black or African American alone | B02001_003E |
| | Asian alone | B02001_005E |
| | Some other race alone | B02001_007E |
| Income | Median household income in the past 12 months (in 2018 inflation-adjusted dollars) | B25119_001E |
| Gender | Male | B01001_002E |
| | Female | B01001_026E |
| Education | Total: Male | B15001_002E |
| | Male, 18 to 24 years, Bachelor's degree | B15001_009E |
| | Total: Female | B15001_043E |
| | Female, 18 to 24 years, Bachelor's degree | B15001_050E |