

U.S. CENSUS BUREAU AND SOCIOECONOMIC STATUS

USING FAMILY SES TO PREDICT STUDENT SUCCESS

Casey Marks

Senior Strategic Data Manager

College of Health and Human Sciences, Purdue University

INAIR Conference 2021. April 8-9, 2021.

Introduction

What is SES? Why is it important?

- Socioeconomic status (SES) is defined as “the social standing or class of an individual or group. It is often measured as a combination of education, income and occupation.” (American Psychological Association)
- Known predictors of student success include High School GPA, Gender, Race/Ethnicity/URM, and Socioeconomic Status (SES), among others
- Immediate application: Evaluation of the effectiveness of the HHS Success Mentors program. Is participation in the Mentor program a significant predictor of student success, when controlling for other known predictors?



Introduction

The Problem

- Purdue's student database lacks a robust measure of SES.
 - Purdue has an indicator of First Generation Student status, which is a binary indicator of parents' education.
 - Direct information on student family income cannot be obtained from the FAFSA, because per federal law, "Without a student's written consent, institutions can only share FAFSA data for the purpose of applying for, awarding, and administering Title IV funds, state aid, and institutional aid programs." (NASFAA, 2019)
 - The same applies to Pell eligibility as to FAFSA information
 - Purdue has no data on students' parents' occupations.
- Question: How do we obtain a reliable measure of SES for use in student success research?

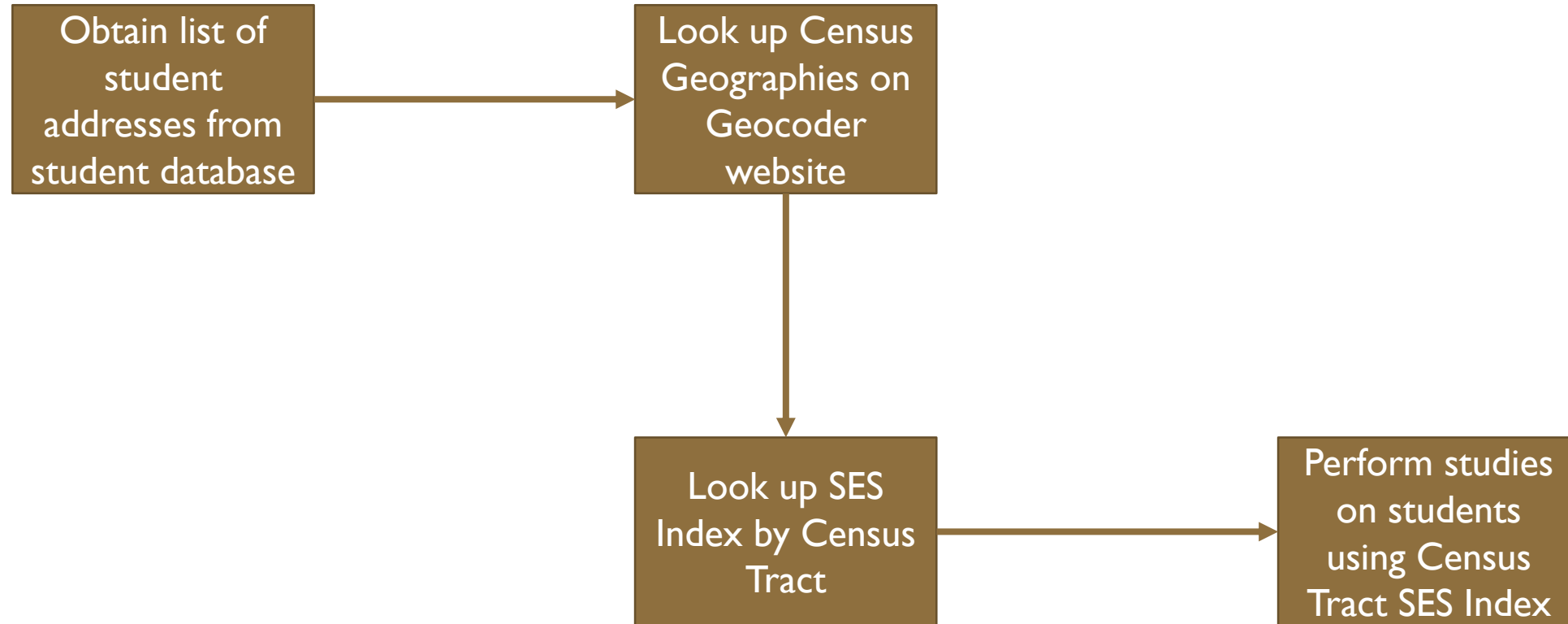
Introduction

Solution to the Problem

- In lieu of direct information about the SES of a student's family, we can estimate the SES of the neighborhood in which the student grew up, using data from the U.S. Census Bureau.

Introduction

Overview of Process



SES Index by Census Tract

Literature Review on the Operational Definition of SES in Higher Education

- Meta-analysis by Rubin (2012a) reviewed 35 papers that had utilized SES in higher education research
- 20 studies used multiple variables to measure SES. I reviewed a sample of 13 of these studies.
- With few exceptions, the studies used a combination of family income, parents' education, and parents' occupations to measure SES.
- Most robust measures used a composite index calculated using an ordinal scale with 6 to 12 levels on each of the following: family income, mother's education, father's education, mother's occupation, father's occupation
- For the studies mentioned above, data was collected from study participants via a survey, which is not a part of standard student institutional data.

SES Index by Census Tract

Yost SES Index (Yost et al, 2001)

- Originally created for use as a control variable in cancer epidemiology research
- Consists of the following 7 factors (Yu et al, 2014):
 - 1) Median household income
 - 2) Median house value
 - 3) Median rent
 - 4) Percent below 150% of poverty line
 - 5) Education Index (Liu et al, 1998)
 - 6) Percent working class
 - 7) Percent unemployed
- Contains factors that account for the three major components of SES: income, education, and occupation. Median house value helps to adjust for regional and local differences in cost of living
- Utilizes information available from the U.S. Census Bureau so that it can be tabulated by Census Tract or Census Block Group



SES Index by Census Tract

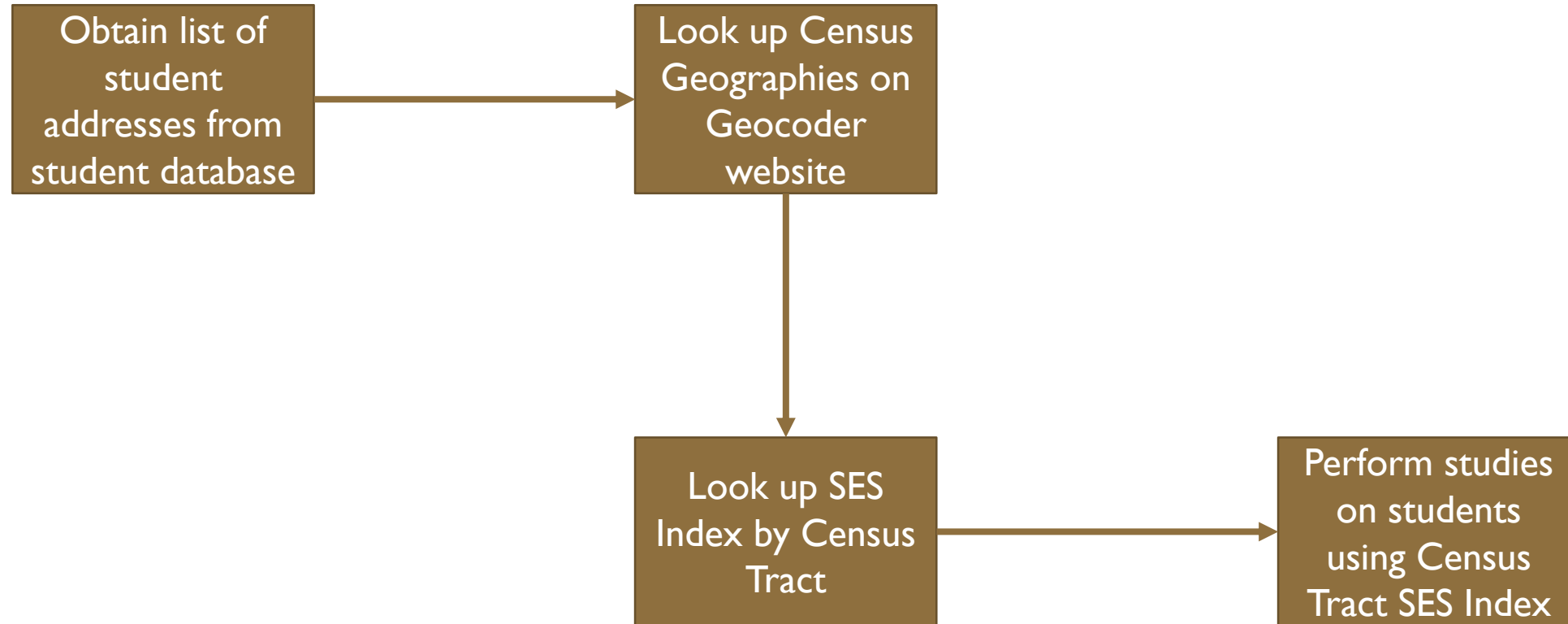
SEER*Stat Database: Yost SES Index by Census Tract

- “Census Tract-level SES and Rurality Database” maintained by the National Cancer Institute’s Surveillance, Epidemiology, and End Results Program (SEER)
- Tabulates the Yost SES Index for every Census Tract in America
- Most recent tabulation uses the American Community Survey (ACS) 5-year data for 2013-2017
- Yost SES Index is available both as a normalized continuous variable and as quintiles
- Database is available to researchers upon request, using the SEER*Stat software
- Login to the database, and download a flat file with Yost CT SES for every Census Tract in the U.S.



Census Geographies

Overview of Process



Census Geographies

Retrieving Geography data using the Census Bureau Geocoder site

- Retrieve a list of students and their parents' addresses from your student database
- Format as a CSV file for input to the Geocoder, with columns as listed in Table 1
- Upload input file to <https://geocoding.geo.census.gov/geocoder/>
 - See next slide for proper selections for a batch file to use with the Yost Census Tract SES database
- Output is a CSV file with many columns, including those listed in Table 1

Table 1: Fields included in input and output files for the Geocoder Geographies batch request

Columns in input CSV file	Columns in output CSV file
Unique ID	Record ID Number
House # and Street Name	Input Address
City	State Code
State	County Code
Zip Code	Tract Code
	Block Code

Census Geographies

Retrieving Geography data using the Census Bureau Geocoder site

The screenshot shows the web browser interface for the Census Bureau Geocoder. The address bar displays `geocoding.geo.census.gov/geocoder/geographies/addressbatch?form`. The page header includes the United States Census Bureau logo and navigation icons. On the left, a navigation menu is visible with the following sections:

- "FIND LOCATIONS USING..." OPTION**
 - One Line
 - Address
 - Address Batch
- "FIND GEOGRAPHIES USING..." OPTION**
 - One Line
 - Address
 - Address Batch** (highlighted with a red box)
 - Geographic Coordinates
- ABOUT DATA...**
 - Benchmarks
 - Vintages

The main content area contains the following form elements:

- Select Address File :** No file chosen
- Benchmark :**
- Vintage :** (highlighted with a red box)
- Batch files may not exceed 10,000 records.
-
- Download a sample CSV file here

The footer contains the following text:

Home | Detailed Information and FAQs | Contact Us

Accessibility | Information Quality | FOIA | Data Protection and Privacy Policy | U.S. Department of Commerce

CONNECT WITH US

Students' Census Tract SES Index

Joining Student Data, Geography Data, and Yost SES Index

- On Geographies list (output from Geocoder site) and Yost SES Index list (from SEER*Stat database), create a key consisting of State Code, County Code, and Census Tract Code.
- Use this key to join Geographies list and Yost SES Index list
- Use Unique ID for a student to join this table to list of students
- Use joined tables to include Yost Census Tract SES Index in list of students
- Have fun doing research on students and SES!

Applications and Results

Regression Model for Student Success

- Dependent variables:
 - 1st Year GPA (Linear Regression)
 - 1-Year Retention (Logistic Regression)
 - 4-Year Graduation (Logistic Regression)

- Independent variables:
 - High School Core GPA
 - Gender
 - Underrepresented Minority (URM)
 - Yost SES Index by Census Tract
 - First-Generation Student
 - Participation as a Mentee in the HHS Success Mentor Program



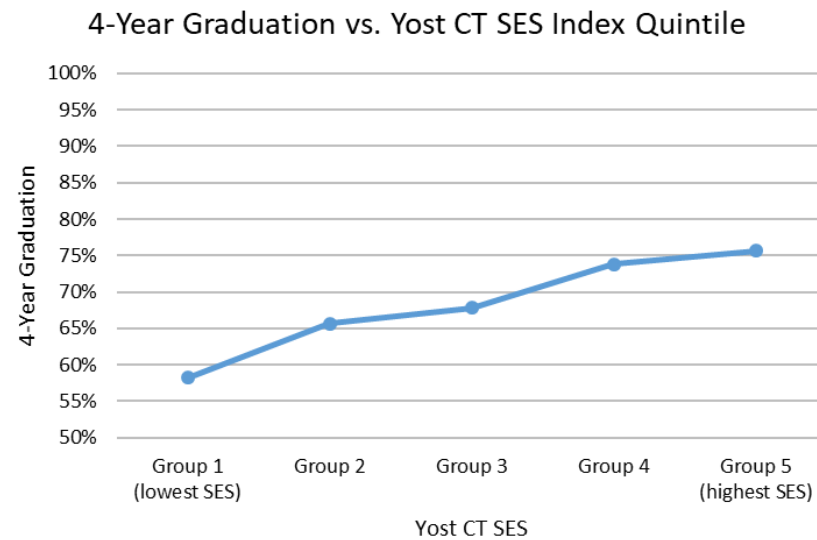
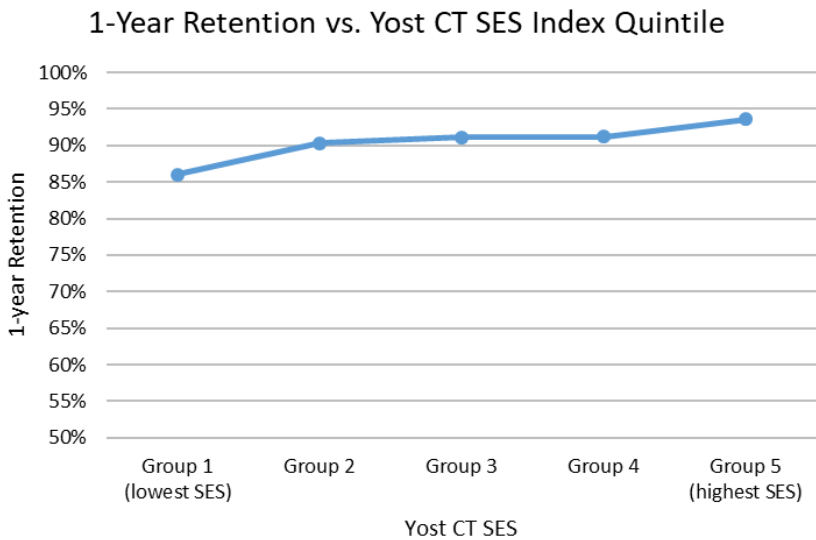
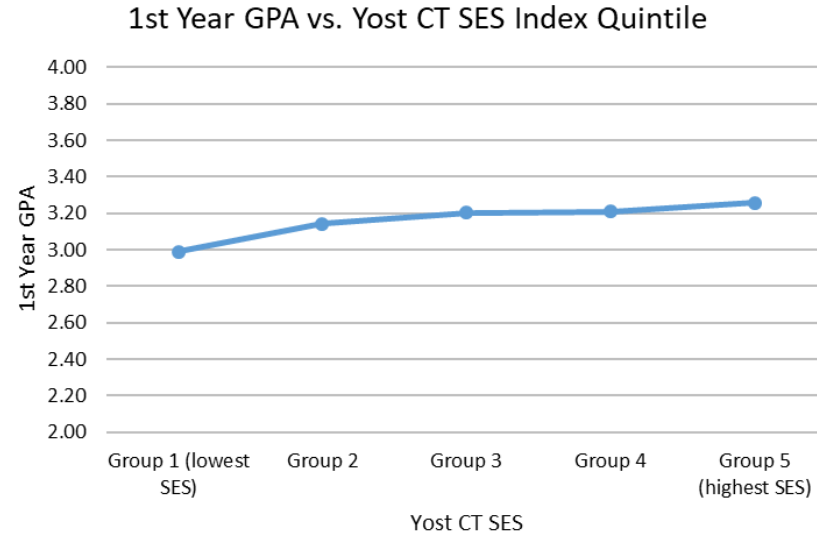
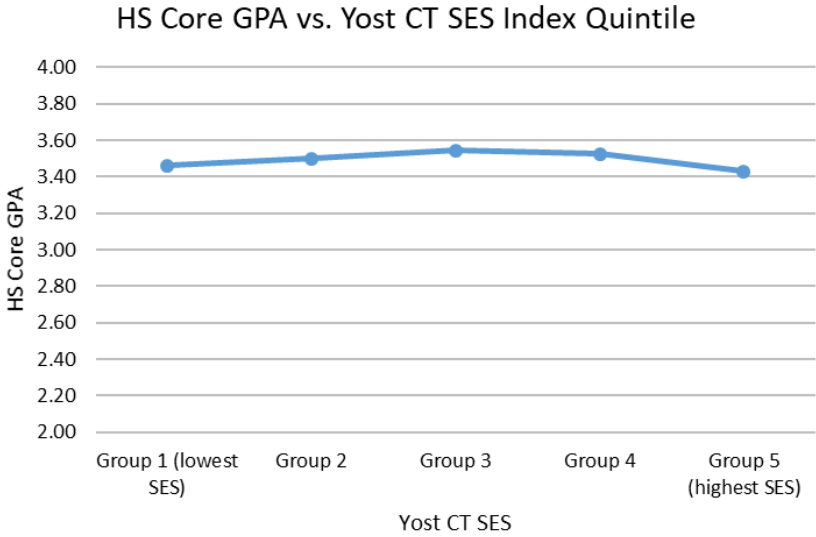
Applications and Results

Regression Model for 1st Year GPA: Results

- Percent of variation in 1st Year GPA accounted for by each significant predictor (R^2):
 - Total R^2 : 35.18%
 - HS Core GPA: 30.00%
 - Yost Census Tract SES: 1.94%
 - First Gen: 1.58%
 - URM: 0.64%
 - Mentee: 0.31%
 - Gender: 0.20%
 - Yost CT SES * URM 0.17%
- Coefficient on Mentee variable: 0.0828 ($p < 0.001$)



Applications and Results



Summary and Conclusions

Concluding thoughts

- Expectation was that low SES students struggle in college, due to difficulties “working the system”, adjusting to the upper-middle-class world of higher education, etc
- Comparisons of quantitative student success measures to SES quintiles confirm expectations

Questions and Answers

- Feel free to contact me at marks11@purdue.edu

Bibliography

American Psychological Association: Psychology Topics: Socioeconomic Status. apa.org/topics/socioeconomic-status

Census Data API User Guide. <https://www.census.gov/content/dam/Census/data/developers/api-user-guide/api-guide.pdf>

Census Tract-level SES and Rurality Database (2000-2015). Surveillance, Epidemiology, and End Results Program. National Cancer Institute. <https://seer.cancer.gov/seerstat/databases/census-tract/index.html>

Financial Aid Data Sharing. NASFAA, 2019. https://www.nasfaa.org/uploads/documents/June_2019_Data_Sharing_White_Paper.pdf

Liu L, Deapen D, Bernstein L. Socioeconomic status and cancers of the female breast and reproductive organs: a comparison across racial/ethnic populations in Los Angeles County, California (United States). *Cancer Causes Control*. 1998 Aug;9(4):369-80.

Purdue Cognos report: Team Content > Departmental Content > Provost > HHS > Enrollment > Enrollment List - Census - w Addresses

Rubin, M. (2012a). Social class differences in social integration among students in higher education: A meta-analysis and recommendations for future research. *Journal of Diversity in Higher Education*, 5, 22-38.

U.S. Census Bureau Geocoder. <https://geocoding.geo.census.gov/>

U.S. Census Bureau Geography Program. <https://www.census.gov/programs-surveys/geography.html>

Understanding and Using American Community Survey Data: What All Data Users Need to Know. https://www.census.gov/content/dam/Census/library/publications/2020/acs/acs_general_handbook_2020.pdf

Yang J, Schupp CW, Harrati A, Clarke C, Keegan THM, Gomez SL. Developing an area-based socioeconomic measure from American Community Survey data. *Cancer Prevention Institute of California, Fremont, California*. 2014.

Yost K, Perkins C, Cohen R, Morris C, Wright W. Socioeconomic status and breast cancer incidence in California for different race/ethnic groups. *Cancer Causes Control*. 2001 Oct;12(8):703-11.

Yu M, Tatalovich Z, Gibson JT, Cronin KA. Using a composite index of socioeconomic status to investigate health disparities while protecting the confidentiality of cancer registry data. *Cancer Causes Control*. 2014 Jan;25(1):81-92.

Appendix

Census Tract vs. Block Group

- “Census tracts generally have a population size between 1,200 and 8,000 people, with an optimum size of 4,000 people... (They) are updated... prior to each decennial census.”
- “Block Groups (BGs) are statistical divisions of census tracts (and) are generally defined to contain between 600 and 3,000 people.” (U.S. Census Bureau Geography Program)
- Pros and Cons: Block Groups are more homogeneous, but they also have higher sampling errors due to their small size
- Yang et al (2014) replicated the Yost SES Index using Census data, comparing data from Tracts with data from Block Groups.
- 81.8% of the Census Tracts and 49.9% of the Block Groups had the highest level of reliability in estimating the SES Index (i.e. all 7 components of the SES Index had a coefficient of variance (CV) < 40%).



Summary and Conclusions

Suggestions for further research

- Use the U.S. Census American Community Survey (ACS) 5-year data to compose a Yost SES Index for every 5-year period in the period of interest for our research, rather than using only one 5-year period for every year of student data
- Explore other SES indices that have been used in social science research that might be adapted for use with U.S. Census Bureau data
- Explore interaction between SES and underrepresented minority (URM) and/or interaction between SES and race/ethnicity in studies on student success



Appendix

API Programming in Python

- I began this project with a plan of writing Python scripts to query the U.S. Census Data API. As the project progressed, I realized that was not the best strategy for what I wanted to do. However, for those who are interested, see below.
- Github repository:
marks11-Purdue/U.S._Census_Bureau_and_Socioeconomic_Status
 - Contains a Python script that looks up geographies in the Geocoder API (“Lookup_Geographies_v21_Final.py”)
 - Uses the Census Vintage that corresponds to the year that a student was a New Beginner
- For API requests in Python, use the “requests” library

