**1/27/2020**

Cross-sectional

Look at school performance – look at test score – school district 🡪 CAASPP 🡪 English version, language arts and math 🡪 cross-sectional or longitudinal🡪 look at 2018,2019 (recent)

San Bernardino school district 🡪 Go to the CAASSP website 🡪 read up on the test 🡪 look at the CAASPP score 🡪 low standard, at standard, above standard 🡪 which specific tests 🡪 language arts and math🡪 variables 🡪 look at gender, race/ethnicity, free and reduce lunch, grades 9th-12 🡪

zipcode(GIS), living particular environment 🡪 school district 🡪 mapping out problematic zipcodes 🡪 layers, crime rates(in zipcode)

Literature review 🡪 zipcode and academic performance, minority population

Spin: Create a spreadsheet that has the citation for the paper we are reading, who the population 🡪 results 🡪 urban or rural, demographic makeup 🡪 do as much as possible 🡪 what makes the population unique

**CAASPP** **– California Assessment of Student Performance and Progress**

Consists of the following assessments:

* Smarter Balanced English language arts/ literacy and mathematics, and science
* California Science Tests
* California Spanish Assessment
* The California Assessment of Student Performance and Progress (CAASPP) System was established on January 1, 2014. The CAASPP System replaced the Standardized Testing and Reporting (STAR) Program, which became inoperative on July 1, 2013
* The CAASPP System encompasses the following assessments and student participation requirements. Information about the content and format of each test is also provided.
* The primary purpose of the CAASPP System is to assist teachers, administrators, students, and parents by promoting high-quality teaching and learning through the use of a variety of assessment approaches and item types.

LITERATURE REVIEW

* SPREADSHEET
* MESS WITH GIS – indicators that are risk factors, high crime factor, food desserts
* PUBLISHED

**2/3/2020**

Look at the specific – social economic - - predominately our population

Look at hypothesis 🡪 where they live 🡪

Environment factors

Culture

Gangs

Community violence

1. Journals 🡪 finish the literature review
2. Do GIS 🡪 argue having parental support is going to trump environmental issues 🡪 affect zipcodes 🡪 parental support and school support 🡪 environmental 🡪 crime rate 🡪 food desert 🡪 liquor 🡪 crucial for a student performance
3. Multiple analysis 🡪 regression 🡪
4. Overlay this on a GIS 🡪 indicators 🡪 put these kids that have support on the map
5. Kids that getting good support 🡪 family
6. Support indicator 🡪 support in school and outside school 🡪 well-rounded support
7. Plop it on a existing GIS indicator
8. High minority majority income students living students 🡪 struggling 🡪 unique population 🡪 kids who have certain county
9. Can we get the GIS map with the indictor 🡪 take all the 2,400 kids 🡪 capability to minipulate the data from CAASP
10. Zipcode, crime rates, food deserts, taking the data and the support🡪 family and school
11. RESTRICTED LOOK AT 5th – 8th graders – middle school, elementary
12. Crucial age period for kids
13. Regression analysis – cross sectional

**2/10/2020**

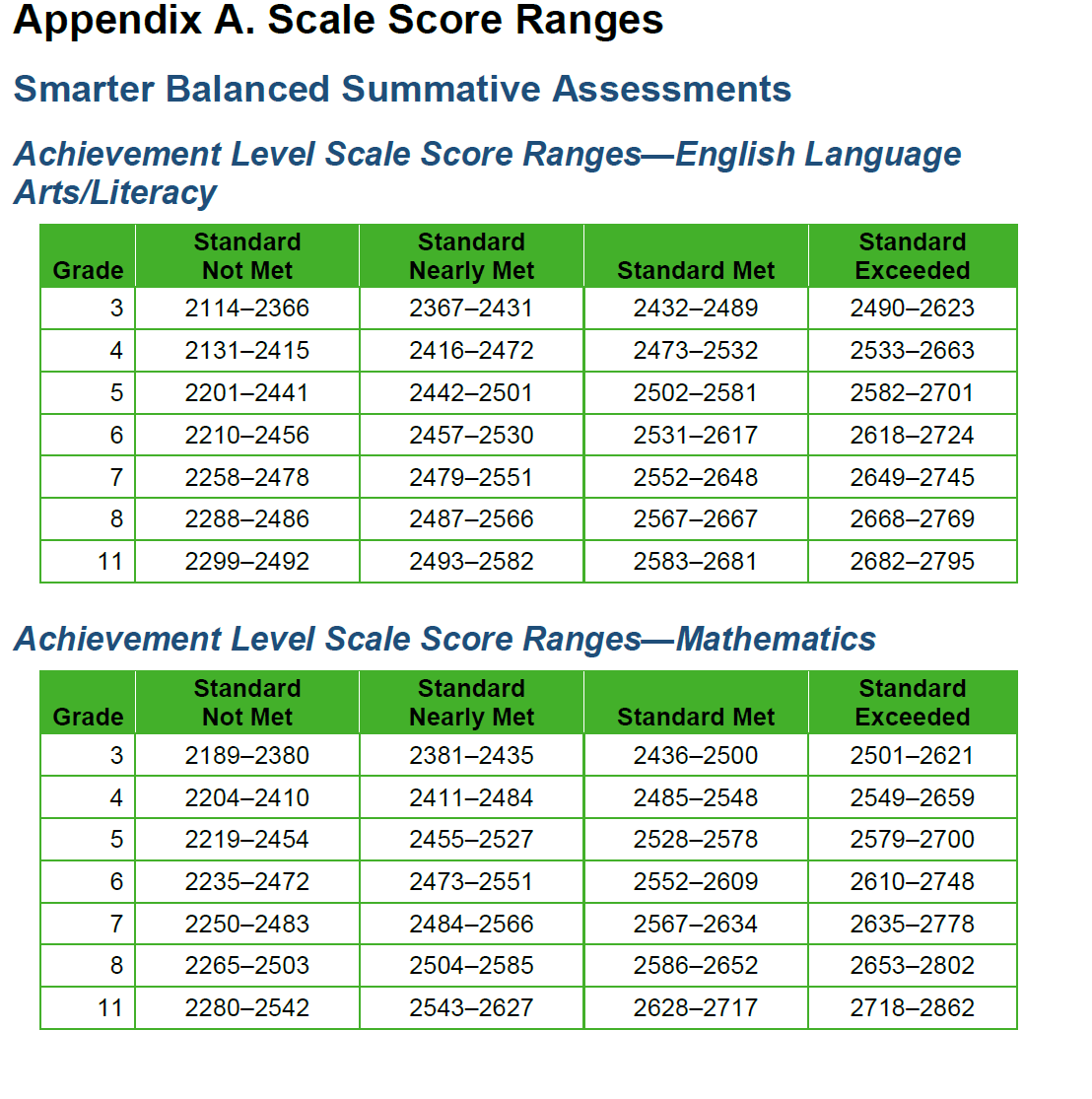
* Define high crime rate, other compare to other crime rates in San Bernardino – specify because….
* Food desert – what is classified – if possible – what percentage of that area
* Limitation – did not have the addresses for the students – recognize
* Find out poverty level rate layer
* Get liquor stores and groceries – see the contrast – not enough groceries stores
* For presentation, multiple pictures – images – contrast
* Multiple questions – running multiple analysis
  + What role does parental support play in student’s test scores
  + Hypothesis: more support equals better grades
  + Get the test score – parental -merge by ID
  + Three different types: support at home, school, support by other adult (binary) – logistic regression for this question, quantify
  + Control for age, gender, grade, race/ethnicity and SES (free and reduced lunch)
  + How do the test scores relate to San Bernardino environmental and social factors
  + Do today, go to the CAASP website and familiar with test – know the background
    - Gender, race/ethnicity 🡪 run the freq for everything - State test
    - The school district measure validity for state and district
    - ELA – English language arts level, they have ELA description
    - ELA claim(?) 🡨
    - ELA sum score
    - If they met the target score and distance score 🡪 how do we overlay the
    - Math 🡨
    - Concern about 🡪 met or exceed that subject language
* **Darrell’s research**
* **Research Question 1**: What role does support play in students test scores?
* *Hypothesis:* more support = better test scores
* *Analysis:* logistic regression
* *Control:* age, gender, grade, race/ethnicity, SES(free and reduced lunch)
* **Research Question 2:** How do test scores vary across spatial social factors in San Bernardino?
* *Hypothesis:* students living in high problematic areas(high crime, food desert, high poverty etc.) will have lower test scores.
* *Analysis:* spatial data analysis
* *Layers:* poverty, crime rates, food deserts
* Have all those variables
* *Note:* So I’m thinking for your overall paper/presentation, it would be beneficial to present research question 2 first and then 1, because question 2 identifies a problem and question 1 identifies a solution.
* **For Next week:**
* **GIS**
* What classifies the hot spots as high crime rate?
* What classifies the areas as food desserts?
* What percentage of the area is considered food dessert
* Poverty level? Explain…
* Try to see if you can find Liquor stores and grocery stores as a layer
* **DATA**
* Understand CAASPP Data
* Run frequencies of CAASPP Variables
* Be familiar with the datasets
* **Limitation**
* Zipcodes can only tell us so much; Lack of addresses is a limitation

SBAC\_ELA\_EXCEEDS\_STANDARDS = pass or fail

Figure out the state average – filter – dataset 🡪 5th to 8th grade

**2/24/2020**

Use achievement score



Still need to find out all the GIS questions – we can’t move without GIS question – additionally – merge data sets – use gender and SES / merging by ID -covariates – use from the school district (CAASP)

Race and ethnicity – 4 categories - HISAPNIC and AFRICAN AMERICANS, white(other minorities -ASIAN,PFI > two categories)

Cleaning up the data – CAASP – create two categories – met the standard and not met standard – math and english

Grades not third or fourth

5,6,7,8, 11

Clean data dob – for age

Calculate response time subtracted by date birth – format – reason why we do that – we want that exact age from dob - control for age , year ,month and date – get rid of response time

**KNOW the definition**

**Week of 02/24**

·      **GIS**

* What classifies the hot spots as high crime rate?

Use use crime Index

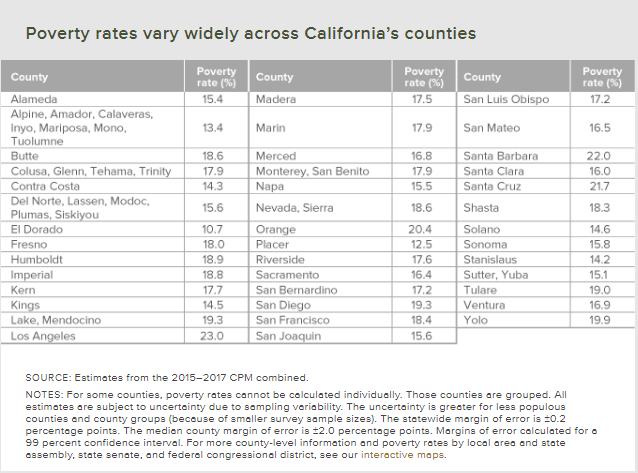
**Fields:**

* + OBJECTID *( type: esriFieldTypeOID , alias: OBJECTID )*
  + Shape *( type: esriFieldTypeGeometry , alias: Shape )*
  + ID *( type: esriFieldTypeString , alias: ID , length: 12 )*
  + NAME *( type: esriFieldTypeString , alias: NAME , length: 13 )*
  + STATE\_NAME *( type: esriFieldTypeString , alias: STATE\_NAME , length: 20 )*
  + ST\_ABBREV *( type: esriFieldTypeString , alias: ST\_ABBREV , length: 2 )*
  + TOTPOP\_CY *( type: esriFieldTypeInteger , alias: 2019 Total Population (Esri) )*
  + TOTPOP\_FY *( type: esriFieldTypeInteger , alias: 2024 Total Population (Esri) )*
  + POPGRWCYFY *( type: esriFieldTypeDouble , alias: 2019-2024 Population: Annual Growth Rate (Esri) )*
  + CRMCYTOTC *( type: esriFieldTypeInteger , alias: 2019 Total Crime Index (AGS) )*
  + CRMCYPERC *( type: esriFieldTypeInteger , alias: 2019 Personal Crime Index (AGS) )*
  + CRMCYMURD *( type: esriFieldTypeInteger , alias: 2019 Murder Index (AGS) )*
  + CRMCYRAPE *( type: esriFieldTypeInteger , alias: 2019 Rape Index (AGS) )*
  + CRMCYROBB *( type: esriFieldTypeInteger , alias: 2019 Robbery Index (AGS) )*
  + CRMCYASST *( type: esriFieldTypeInteger , alias: 2019 Assault Index (AGS) )*
  + CRMCYPROC *( type: esriFieldTypeInteger , alias: 2019 Property Crime Index (AGS) )*
  + CRMCYBURG *( type: esriFieldTypeInteger , alias: 2019 Burglary Index (AGS) )*
  + CRMCYLARC *( type: esriFieldTypeInteger , alias: 2019 Larceny Index (AGS) )*
  + CRMCYMVEH *( type: esriFieldTypeInteger , alias: 2019 Motor Vehicle Theft Index (AGS) )*
  + TOTCAGG *( type: esriFieldTypeDouble , alias: 2019 Total Crime Aggregate (AGS) )*
  + PERCAGG *( type: esriFieldTypeDouble , alias: 2019 Personal Crime Aggregate (AGS) )*
  + MURDAGG *( type: esriFieldTypeDouble , alias: 2019 Murder Aggregate (AGS) )*
  + RAPEAGG *( type: esriFieldTypeDouble , alias: 2019 Rape Aggregate (AGS) )*
  + ROBBAGG *( type: esriFieldTypeDouble , alias: 2019 Robbery Aggregate (AGS) )*
  + ASSTAGG *( type: esriFieldTypeDouble , alias: 2019 Assault Aggregate (AGS) )*
  + PROCAGG *( type: esriFieldTypeDouble , alias: 2019 Property Crime Aggregate (AGS) )*
  + BURGAGG *( type: esriFieldTypeDouble , alias: 2019 Burglary Aggregate (AGS) )*
  + LARCAGG *( type: esriFieldTypeDouble , alias: 2019 Larceny Aggregate (AGS) )*
  + MVEHAGG *( type: esriFieldTypeDouble , alias: 2019 Motor Vehicle Theft Aggregate (AGS) )*
  + Shape\_Length *( type: esriFieldTypeDouble , alias: Shape\_Length )*
  + Shape\_Area *( type: esriFieldTypeDouble , alias: Shape\_Area )*

o   What classifies the areas as food desserts?

o   What percentage of the area is considered food dessert

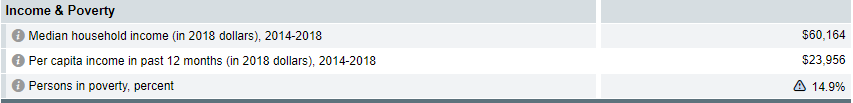
o Poverty level? Explain…



The GIS map shows poverty rates, not income levels. The poverty rate is the ratio of the number of people (in a given age group) whose income falls below the poverty line; taken as half the median household income of the total population.

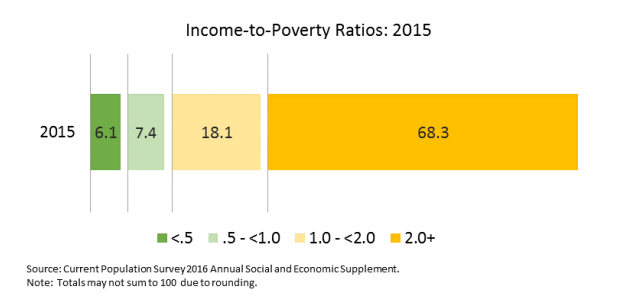


<https://aspe.hhs.gov/2019-poverty-guidelines>



<https://www.census.gov/quickfacts/sanbernardinocountycalifornia>

<https://www.census.gov/newsroom/blogs/random-samplings/2018/09/varying_degrees_ofp.html>



Try to see if you can find Liquor stores and grocery stores as a layer

**DATA CLEANING**

·      **Merge** CAASPP data with other data set

o   We will use gender, low income status, race/ethnicity from School district data

o   Race/ethnicity categories should be: Hispanic, African American white, other minorities

·      For the Ela and math level, categorize into “met the standard”(3 and 4) and “not met the standard” (1 and 2) 🡪 1 = true , 2 = 0

·      Run frequencies for variables in the merged data.

o   Age (day of survey – date of birth)

o   Race/ethnicity

o   Gender

o   Low-income status

**3/2/2020**

**Week of March 2, 2020**

**GIS**

Continue to make layers for grocery stores

Rewrite the explanations for GIS that disappeared.

R Analysis

·      Table 1

o   AGE

o   GENDER

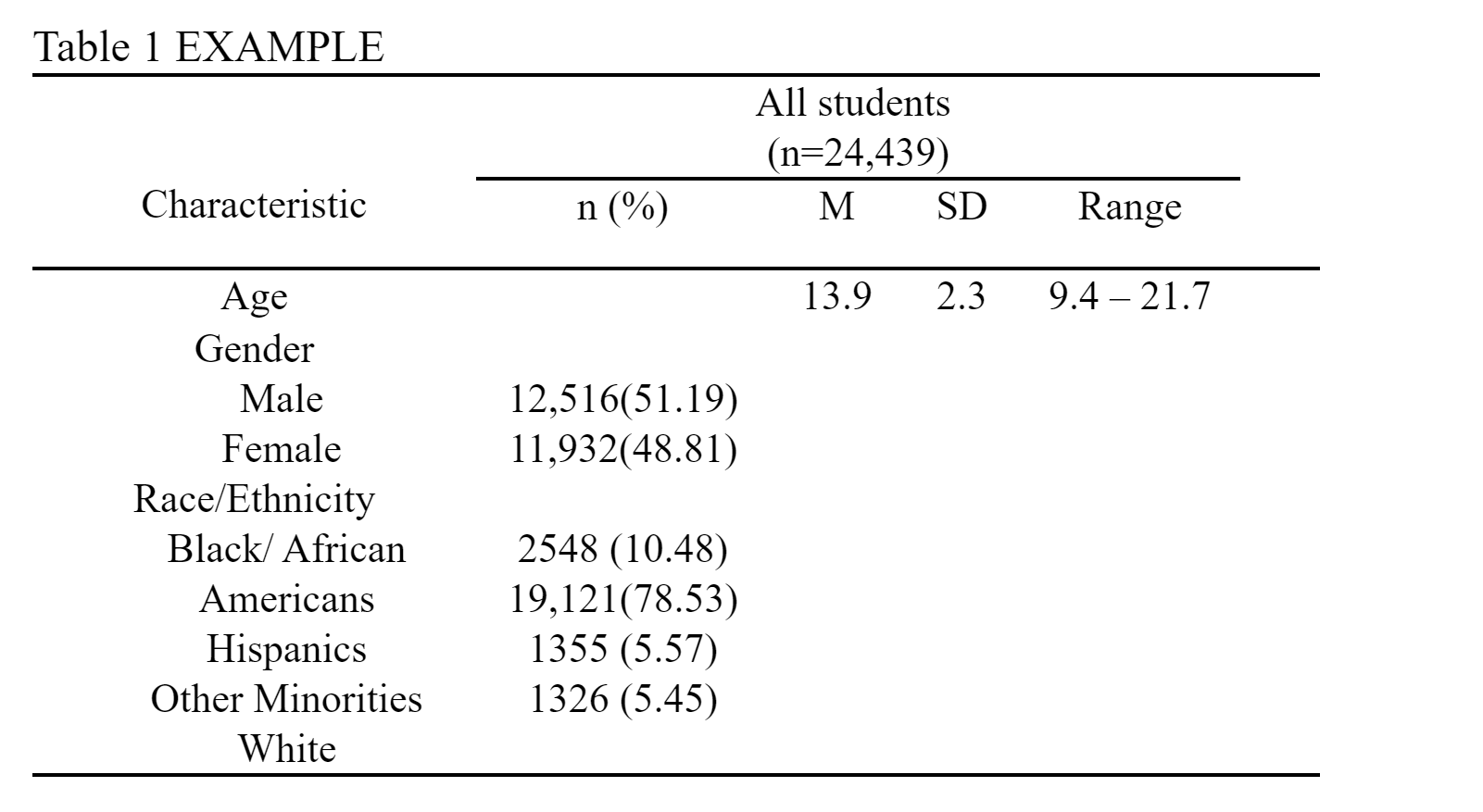
o   GRADE

o   RACE/ETHNICITY

o   LOW SES

o   ELA

o   MATH



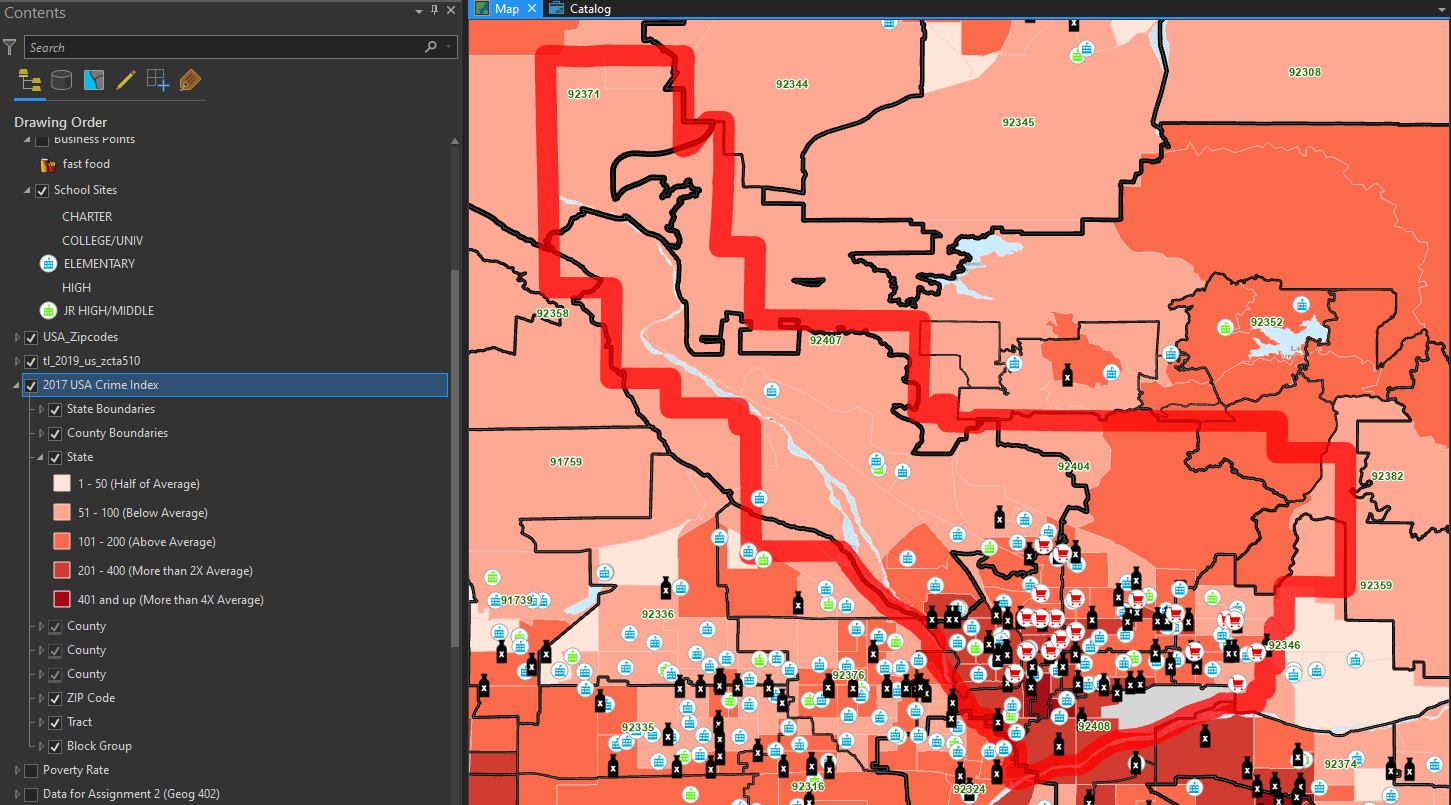
Forget about the logistic 🡪 know the main exposure based on exposure 🡪 physical variables 🡪 look at individual variable or composite 🡪 modify

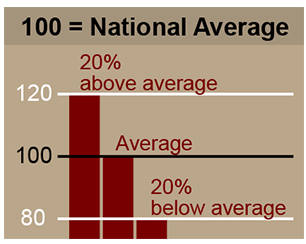
Table 1 EXAMPLE\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | All Students  (n = 16,422) | | CAASP  English | | CAASP  Math | |
| Characteristics | n (%) | M | SD | Range | Standard Met  n (%) | Standard Not Met  n (%) | Standard Met  n (%) | Standard Not Met  n (%) |
| Age |  | 13.29 | 2.18 | 9.8 – 19.6 |  |  |  |  |
| Gender |  |  |  |  |  |  |  |  |
| Male | 8383 (51.05) |  |  |  | 3,178 (44.48) | 5,270 (56.13) | 2,152 (50.41) | 6,201 (51.25) |
| Female | 8039 (48.95) |  |  |  | 3,996 (55.52) | 4,041 (43.87) | 2117 (49.59) | 5,899 (48.75) |
| Race/Ethnicity |  |  |  |  |  |  |  |  |
| African Americans | 1488 (9.10) |  |  |  | 518 (34.81) | 970 (65.19) | 235 (15.82) | 1,250 (84.18) |
| Hispanics | 12,972 (79.32) |  |  |  | 5,647 (43.53) | 7,325 (56.47) | 3,337 (25.69) | 9,650 (74.31) |
| Other Minorities | 1,067 (6.52) |  |  |  | 510 (47.80) | 557 (52.20) | 371 (34.74) | 697 (65.26) |
| Whites | 828 (5.06) |  |  |  | 469 (56.64) | 359 (43.36) | 326 (39.32) | 503 (60.68) |
| Low SES status |  |  |  |  |  |  |  |  |
| Yes | 13,952 (85.31) |  |  |  | 5,803 (41.59) | 8,149 (58.41) | 3,447 (24.69) | 10,515 (75.31) |
| No | 2403 (55.81) |  |  |  | 1,341 (55.81) | 1,062 (44.19) | 822 (34.15) | 1,585 (65.85) |
| Grade level |  |  |  |  |  |  |  |  |
| 4 | 2,823 (17.26) |  |  |  | 1,084 (38.40) | 1739 (61.60) | 859 (30.36) | 1,970 (69.64) |
| 5 | 3,050 (18.65) |  |  |  | 1,243 (40.75) | 1,807 (59.25) | 649 (21.25) | 2,405 (78.75) |
| 6 | 2,957 (18.08) |  |  |  | 1,274 (43.08) | 1,683 (56.92) | 847 (28.62) | 2,112 (71.38) |
| 7 | 2,761 (16.88) |  |  |  | 1,250 (45.27) | 1,511 (54.73) | 760 (27.50) | 2,004 (72.50) |
| 8 | 2,498 (15.27) |  |  |  | 1,106 (44.28) | 1,392 (55.72) | 682 (27.32) | 1,814 (72.68) |
| 11 | 2,266 (13.86) |  |  |  | 1,187 (52.38) | 1,079 (47.62) | 472 (20.82) | 1,795 (79.18) |
| All Students | 16,422 |  |  |  | 7,144 (43.68) | 9,211 (56.31) | 4,269 (26.08) | 12,100 (73.92) |

GIS Explanation

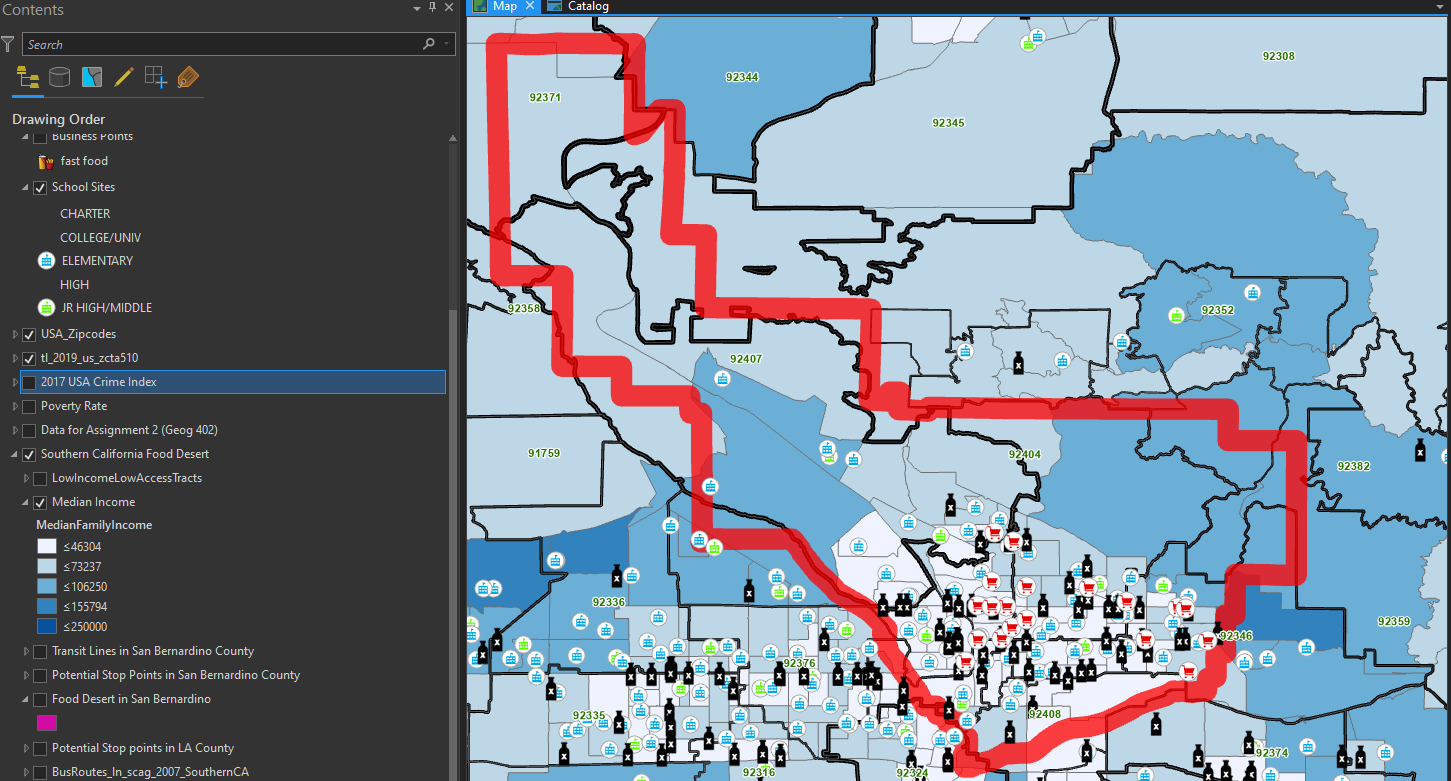
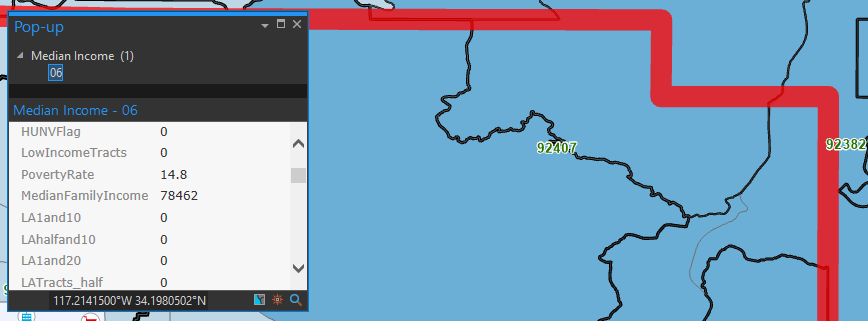
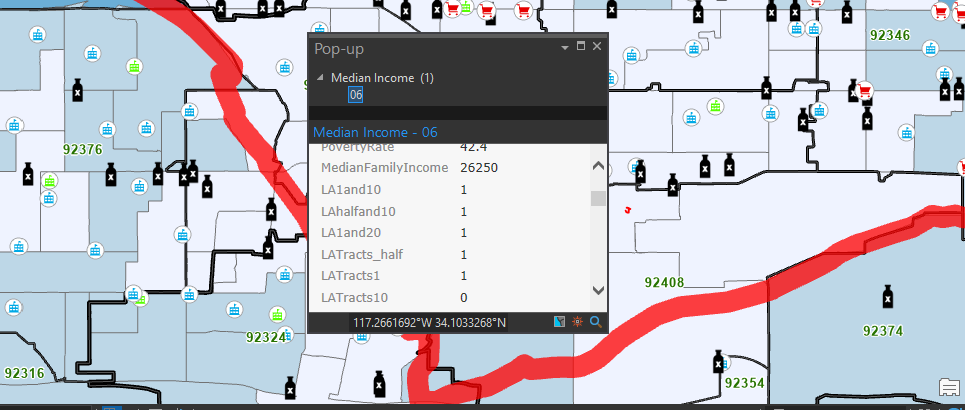
Crime Index Explanation: GIS data included

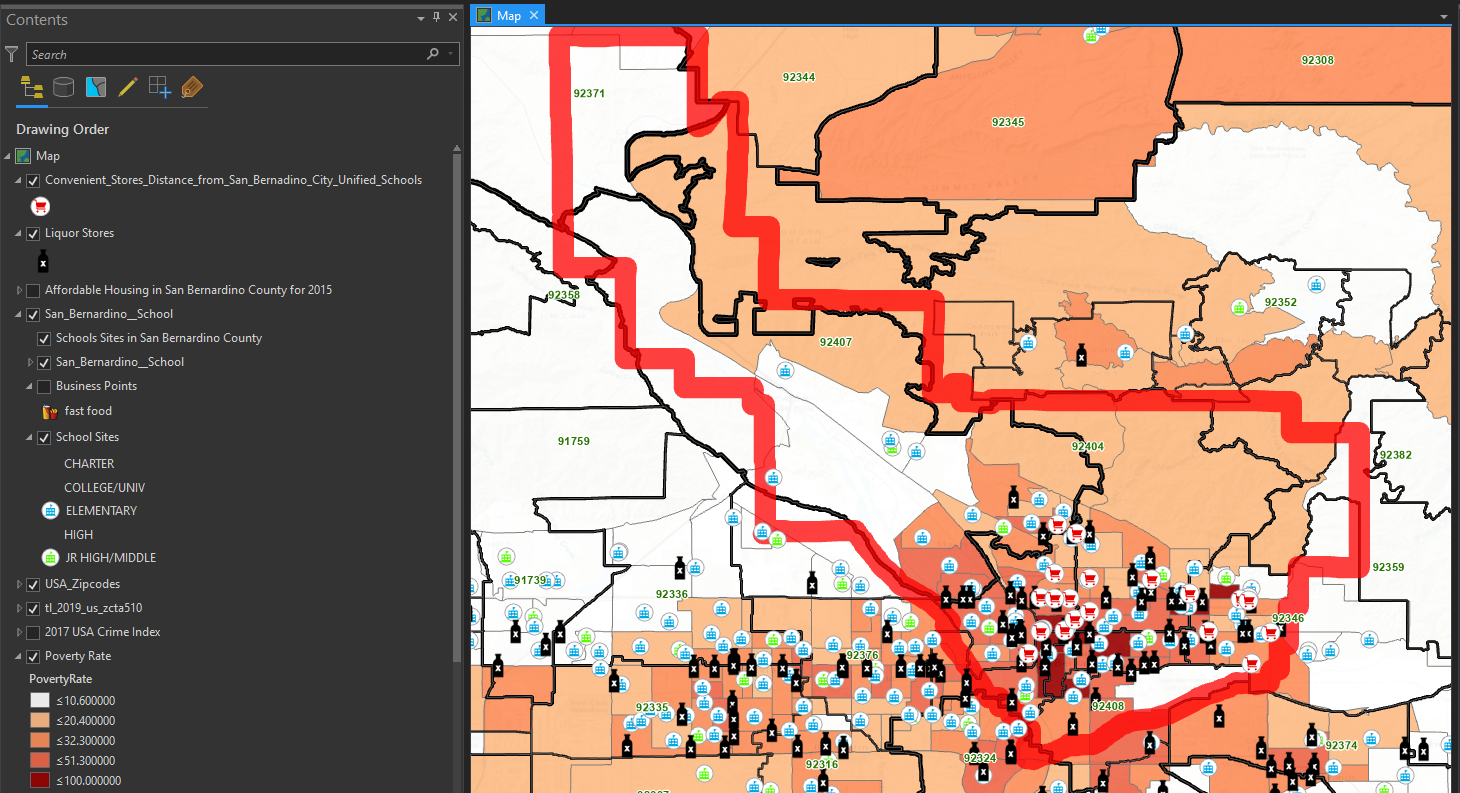
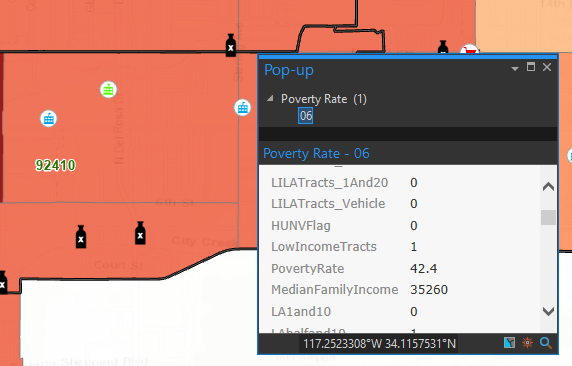
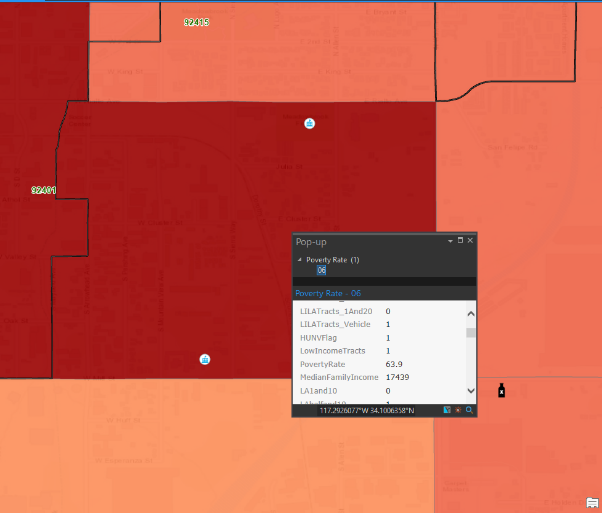
* Total Crime Index
* Personal Crime Index
* Murder Index
* Rape Index
* Robbery Index
* Assault Index
* Property Crime Index
* Burglary Index Larceny Index
* Motor Vehicle
* Theft Index



The values are all referenced by an index value. The index values for the US level are 100, representing average crime for the country. A value of more than 100 represents higher crime than the national average, and a value of less than 100 represents lower crime than the national average. For example, an index of 120 implies that crime in the area is 20 percent higher than the US average; an index of 80 implies that crime is 20 percent lower than the US average.

Median Income Explanation



Poverty Rate Layer Explained

The **poverty rate** is the percentage of people who are living in **poverty**, meaning that they are making below the amount of money deemed necessary by the federal government to have a basic standard of living

**Still working on grocery store layer**

3/9/2020

We look at crime rates at, ask dr. wiafe for questions - **aggregate** –

Food desert- farther away than 1 mile from a supermarket

½ mile from grocery store, you have access

Variables that determine support – measures within a support –

Support:

The protective factors assessed included:  support at home, support at school, support outside home and school, sleep, and growth mindset. Support at home was assessed by 3 questions pulled from the *California Healthy Kids Survey Resiliency Module(25)*.  The questions asked, “in my home, there is a parent or some other adult…a)who is interested in my school work, b)who talks with me about my problems, c)who listens to me when I have something to say”. Support outside the home and school was also retrieved from the *California Healthy Kids Survey Resiliency Module(25)* and it asked, “outside of my home and school, there is an adult…a)who really cares about me, b)who notices when I am upset about something, c) whom I trust”. In response, participants could select, “Not at all true”, “A little true”, “Pretty much true”, and “very much true” for both these factors.

Support at school was measured by how students responded to the following statements: “adults at this school encourage me to work hard so I can be successful in college or at the job I choose”, “My teachers work hard to help me with my schoolwork when I need it”, “teachers give students a chance to take part in classroom discussions or activities”, “this school promotes academic success for all students”, and “teachers go out of their way to help students”. Responses were on a five-point Likert scale, from 1=strongly agree to 5=strongly disagree.

Here's a story – income 🡪 food desert 🡪 crime index 🡪 dump school 🡪 pause 🡪 new layer 🡪 take the schools out 🡪 plot our students 🡪 given their income, poverty 🡪 food desert

Poverty

Food Desert

Crime Index

Morning,

Darrelle, good job on your work so far. we have made a ton of progress this quarter and I am so excited to pull it all together next quarter.

Do not stress about the work below. I want you to focus on your finals and to actually rest over spring break.  So do as much as possible.

**Week of March 9 – 30th, 2020**

**GIS**

·      What does “100” mean as the national average for crime rate? How many crimes is that?

·      Give me a tally of income?

**Data Cleaning (look at the write up I gave you this morning to guide this)**

Support:  
  
The protective factors assessed included:  support at home, support at school, support outside home and school, sleep, and growth mindset. Support at home was assessed by 3 questions pulled from the California Healthy Kids Survey Resiliency Module(25).

The questions asked, “in my home, there is a parent or some other adult…

a)who is interested in my school work,

b)who talks with me about my problems,

c)who listens to me when I have something to say”.

Support outside the home and school was also retrieved from the California Healthy Kids Survey Resiliency Module(25) and it asked, “outside of my home and school, there is an adult…

a)who really cares about me,

b)who notices when I am upset about something,

c) whom I trust”. In response, participants could select, “Not at all true”, “A little true”, “Pretty much true”, and “very much true” for both these factors.

Support at school was measured by how students responded to the following statements: “adults at this school encourage me to work hard so I can be successful in college or at the job I choose”, “My teachers work hard to help me with my schoolwork when I need it”, “teachers give students a chance to take part in classroom discussions or activities**”, “this school promotes academic success for all students”**, and “teachers go out of their way to help students”. Responses were on a five-point Likert scale, from 1=strongly agree to 5=strongly disagree. **Cant find that variable in the dataset**

**Recoding**

*Support at home* and *support outside home and school* (in the large school data set)

A little true, not at all true = 0

Pretty much true and very much true =1

·      *Support at school* (school climate data set)

o   Strongly disagree, disagree and neither = 0

o   Agree and strongly agree  = 1

**Table 1**

·      Add the different types of support on to Table 1

Combine data sets

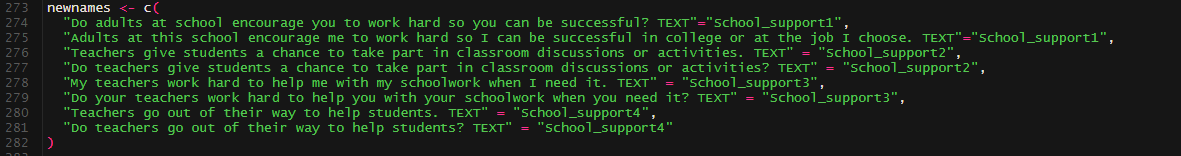
Combined + sbcused+CORE 2018-2019 elementary = Combined1

Combined1+sbcusd+CORE 2018-2019 secondary = Combined2

Did not work so we have to

must do more data cleaning, created datasets climateELE and climateSEC

, noticed each one had different sounding questions - make one variable to similar sounding questions

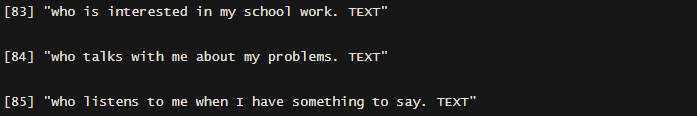


“in my home, there is a parent or some other adult…

a)who is interested in my school work,

b)who talks with me about my problems,

c)who listens to me when I have something to say”.



Support outside the home and school was also retrieved from the California Healthy Kids Survey Resiliency Module(25) and it asked, “outside of my home and school, there is an adult…

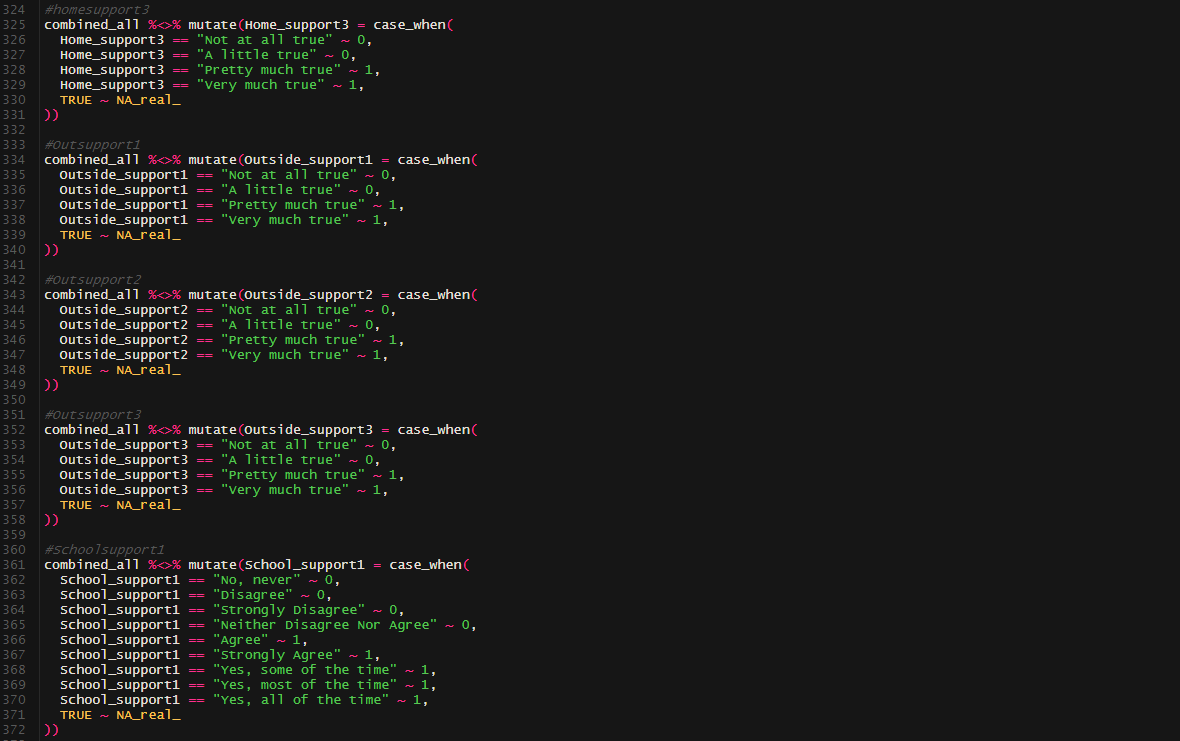
a)who really cares about me,

b)who notices when I am upset about something,

c) whom I trust”. In response, participants could select, “Not at all true”, “A little true”, “Pretty much true”, and “very much true” for both these factors



**CODING**



**Add new data to table1**

**GIS questions**

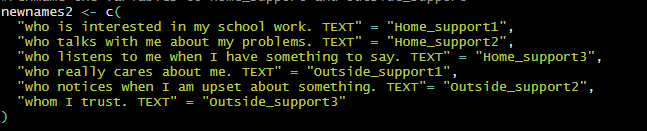
·      What does “100” mean as the national average for crime rate? How many crimes is that?

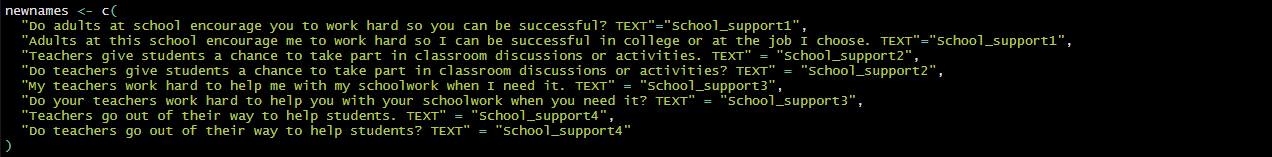
·      Give me a tally of income?

**Research Journal 3/29/2020**

**Merging datasets of Combined( included merged CAASP and SB School survey) and ClimateELE (elementary climate data) and ClimateSEC (Secondary climate data).**

**The problems at first with merging ClimateELE and ClimateSEC was the same but questioned differently in each survey. For example, in ClimateELE survey, “Do adults at school encourage you to work hard so you an be successful? TEXT” was questioned differently in ClimateSEC survey as, “Adults at this school encourage me to work hard so I can be successful in college or at the job I choose. TEXT.” To solve this problem, I decided to categorize the questions from each dataset as, “School\_support1”, School\_support2”, and so forth according to the questions being asked and matched.**

**Below contains the coding of what I did:**



**Support was based on questions based on**

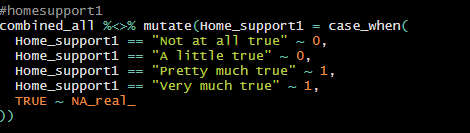
* **Home\_support (support at home)**
* **Outside\_support (outside adult from school and home)**
* **School \_support (support at school)**

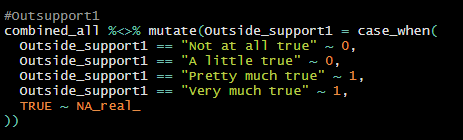
**However I was not able to find the question, “ this school promotes academic success for all students in either datasets.**

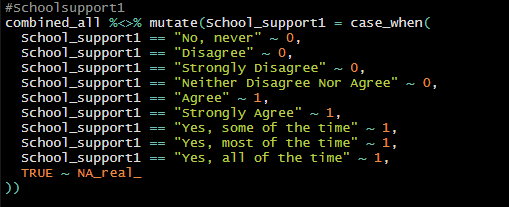
**To make sense of the coding**

* **Home\_support1: “Who is interested in my school work”**
* **Home\_support2: “Who takes with me about my problems”**
* **Home\_support3: “Who listens to me when I have something to say”**
* **Outside\_support1: “Who really cares about me.”**
* **Outside\_support2: “Who notices when I am upset about something”**
* **Outside\_support3: “Whom I trust”**
* **School\_support1: “Do adults at school encourage you to work hard so you can be successful**
* **School\_support2: “ Teachers give students a chance to take part in classroom discussions**
* **School\_support3: “Teachers work hard to help me with my schoolwork when I need it”**
* **School\_support4: “Teacher’s go out of their way to help students”**

**In terms of coding or recoding variables,**







**For Home\_support and Outside\_support had the same given answers. So coding from character to numeric**

* “Not at all, true”, “A little true”, “Pretty much true”, “Very much true” = 0,
* “Pretty much true”, “Very much true” = 1

**For School\_support1**

* “No, never”, “Disagree”, “Strongly Disagree”, Neither Disagree Nor Agree” = 0
* “Agree”, Strongly Agree”, “Yes, some of the time, “Yes, most of the time”, “ Yes, all of the time” =1

**I sent two separate datasets which are the ones listed below. They are basically the same thing, the only thing different is that Cleaned\_Combined\_all contains only the necessary variables we are using for analysis so it contains less of the column variables. Combined\_all contains all the variables.**

|  |  |
| --- | --- |
| Datasets | |
| Combined\_all | Contains all variables from CAASP, SBSURVEY,CLIMATEELE,CLIMATESEC |
| Cleaned\_Combined\_all | Contains only necessary variables for analysis:  ID, zipcode, ageatstudy, ethnicity, lowses status, ela\_level,math\_level, Home\_support#, Outside\_support#,School\_support# |

**The other file I sent is the “Table 1 and Table 2” file which contains the demographics and support tables. We cover the demographic table a couple weeks ago, but I created a new table for student support related to passing CAASP English and Math. Below is the description of what I see for each variable for the table.**

**We see that most variables showed that there was a higher frequency of students passing with support compared to students with no support in both math and English according to variables of students who had support vs no support.**

**Home\_support1,Home\_support2, Home\_support3, Outside\_support1, Outside\_support3, School\_support1, School\_support2**

**However, in both cases with support and no support, there was higher number of students who did meet the Standard. Some cases even showed that students that had no support did better than students with supports has shown with Outside\_support2, School\_support3, and School\_support4.**

**3/31/2020**

**Analysis – Support categorizes**

**Aggregate – three into one category 🡪 all home support 1 category – outside support 1 category, school support category**

**Logistic regression 🡪 English and Math aggregate**

**Home support + Outside support + School support**

**OR**

**Confidence interval**

**Reference categories – explanation – figure**

**Choose why reference category**

**GIS questions**

·      What does “100” mean as the national average for crime rate? How many crimes is that?

·      Give me a tally of income?

Darelle,

Good Job on all your work thus far! It looks great.

I really appreciate you being on top of things and going above and beyond in your work. It absolutely makes things easy and makes moving along on this research project feasible.

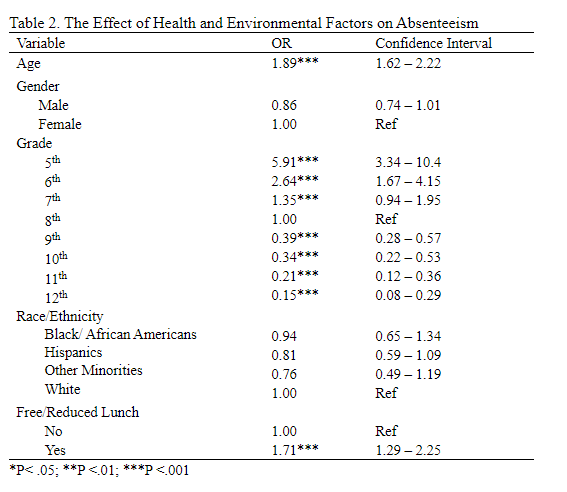
For this week please work on the following.

**Week of March 30th, 2020**

**Run Logistic Regressions and make a table with the results.**

**English** = age + gender + race/ethnicity + SES + Grade level + Support at home + support at school + support outside home and school

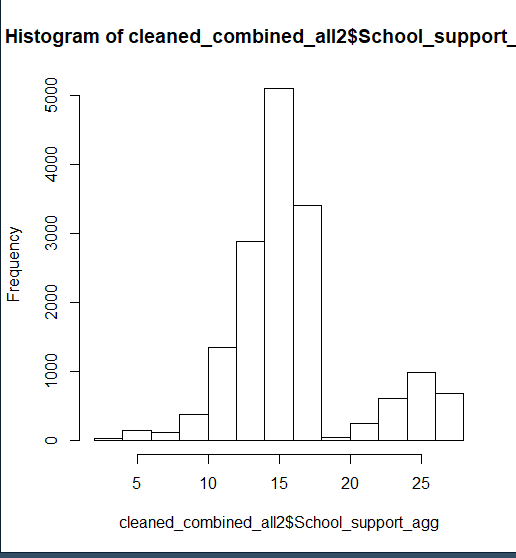
**Math** = age + gender + race/ethnicity + SES + Grade level + Support at home + Support at school + Support outside home and school



**4/1/2020 Dr.Shavlik**

1. Code independent variables from 1 to 4  
   2. Sum the three variables to single variable. E.g. Home=home1+home2+home3. This should have values ranging from 3 to 12  
   3. Keep 4 level outcome variable in final file **(?)** 🡪 categorizing 4 levels   
   4. Write R code the reads final file in directly and then does logistic regression on it.  
   5. Darelle listed his hypotheses in the Research Journal

**Take the aggregate column, each person going to have a value**



**Difference between 5 through 8th grade**

**What is the distribution of the school ids between 5th graders and 4th graders**

**What is the correlation that create the difference**

> cleaned\_combined\_all2 %>% filter(stu\_grade==5) -> d

> hist(d$School\_support\_agg)

> cleaned\_combined\_all2 %>% filter(stu\_grade==4) -> d

> hist(d$School\_support\_agg)

> cleaned\_combined\_all2 %>% filter(stu\_grade==7) -> d

> hist(d$School\_support\_agg)

> cleaned\_combined\_all2 %>% filter(stu\_grade==8) -> d

> hist(d$School\_support\_agg)

> cleaned\_combined\_all2 %>% filter(stu\_grade==6) -> d

> hist(d$School\_support\_agg)

> cleaned\_combined\_all2 %>% filter(stu\_grade==11) -> d

> hist(d$School\_support\_agg)

> hist(cleaned\_combined\_all2$School\_support\_agg)

> hist(cleaned\_combined\_all2$School\_support\_agg)

> cleaned\_combined\_all2 %>% filter(stu\_grade==4) -> d

> table(d$School\_support\_agg)

**Find what’s causing the difference 🡪 Create a binary**

**4th graders are smallest peak**

**Ask for the original questionnaire from Nipher 🡪 original measurement**

**Maybe exclude 4th graders from dataset**

**Numeric Categorization 🡪 quartiles**

**Because student grade level ->**

**ELE 🡪 5th grade**

**Secondard 🡪 everything else**

**Mismatch in the questions 🡪 error I the coding in the reponses for secondary and elementary school**

**Week of April 6th, 2020**

Trouble shoot any issues on the logistic regression

Start writing Methods Section(see attachment for guidance). It should include the following:

-       Participants and Procedures

-       Measures

-       Data Analysis

Start writing a Results Section. It should include the following:

-       A Paragraph or 2 on demographics

-       Regression analysis paragraphs (2+ paragraphs)

GIS

TASKS TO DO

* Create level 4 factors
* Create new descriptive table with ordinal variables
* Create new descriptive table using cleaned\_combined\_all3 for consistency
* Analysis on ordinal variables
* Consulting with R-Professional biostatistician on potential problems with logistic regression modeling on R – most CI intervals were negative results. – Changing ref to group to 11 for grades 🡪 changes OR
* Work on GIS
* Write a **Methods section** 🡪 Using R, write it down
* Results 🡪 Basic demographics
* Figure out how the States collects CAASP data
* Publication table – easy to read – consult with Dr.Shavlik reasons for publication 🡪 Standard nearly met 🡪 still means you failed the class. Either you pass or you failed.

**4/8/2020**

**Proc logistics 🡪 SAS 🡪 ask Paul**

**4 variables 🡪losing power 🡪 ask Nipher, turning things 🡪 why not categorize them and just add them together 🡪 collapse it at the end. Losing information 🡪 categorize after aggregating data 🡪 Support**

**Categorized versions of ordinal variables 🡪 run categorize variables**

**Outside and home**

**Run 2 models 🡪 4 and 5 categories**

**Collapse**

**11 and 12 🡪 4**

**9 and 10 🡪 3**

**7 and 8 🡪 2**

**6 and below 🡪 1 reference**

**School Support**

**16 and 15 🡪 5**

**14 and 13 🡪 4**

**12 and 11 🡪 3**

**10 and 9 🡪 2**

**8 and 4 🡪 1**

**Figure out analytic sample size**

**2 models**

**Question is 🡪 how many people are missing**

**Sample size of the first model**

**Sample size of the second model**

**Make the decision of analytic sample size**

**You have to exclude the missing in the model**

**Have to have a where statement to exclude the missing variables so to have a consistent sample size**

**Challenge people 🡪 understand the why**

**4/12/2020**

*Meeting updates*:

It ok to use the binary version of Pass/Failed (1/0) categories of sbac\_math\_level2cat and sbac\_ela\_level2cat. When running the frequencies of the ordinal versions of sbac\_math\_level and sbac\_ela\_level, it made sense to collapse. **Such we will use sbac\_math\_level2 and sbac\_ela\_level2cat for our outcome**.

There are two versions of Support\_Agg (Home,Outside, and School) which are the binary version (collapsed levels to 0 and 1) and the factor level 4 version( 1-4). He said to add the factor level 4 version then collapsed it after aggregating it. Such for example:

Outside and home

Run 2 models 🡪 4 and 5 categories

Collapse

11 and 12 🡪 4

9 and 10 🡪 3

7 and 8 🡪 2

6 and below 🡪 1 reference

School Support

16 and 15 🡪 5

14 and 13 🡪 4

12 and 11 🡪 3

10 and 9 🡪 2

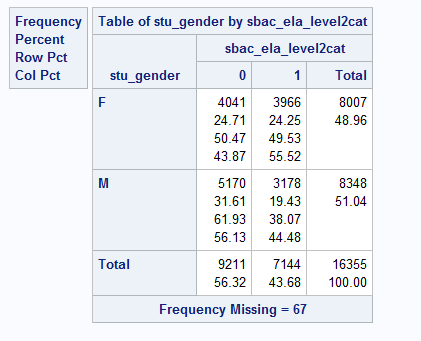
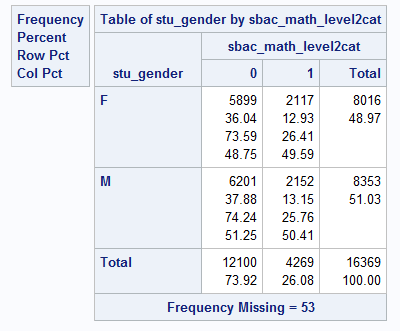
8 and 4 🡪 1

Name these variables as \_supportcl and compare with the \_support\_agg.

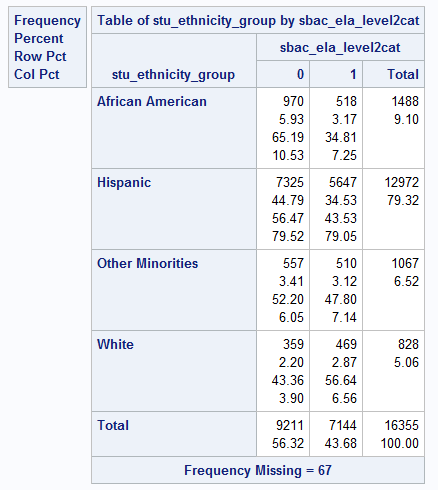
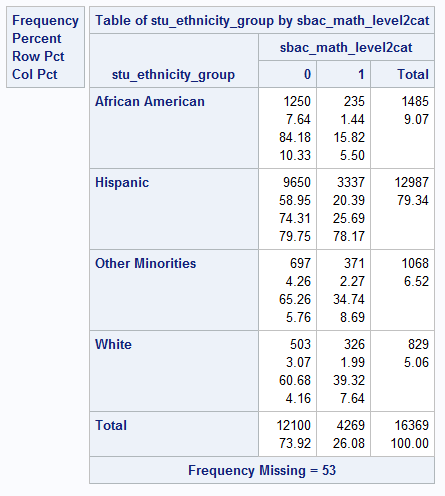
Using SAS for analysis so Dr.Shavlik can see the numbers clearly. Will use both R and SAS to compare analysis. These will be referred as model 2 in my work journal below.

*Work:*

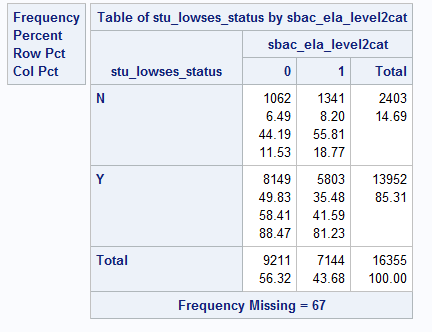
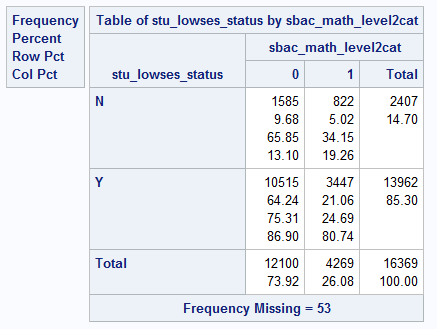
Frequencies for Gender and outcomes (ELA and Math) using SAS on cleaned\_combined\_all3 dataset



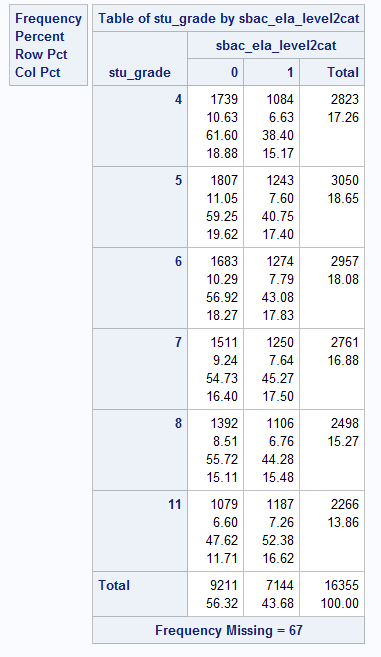
Frequencies for Ethnicity and outcomes (ELA and Math) using SAS on cleaned\_combined\_all3 dataset

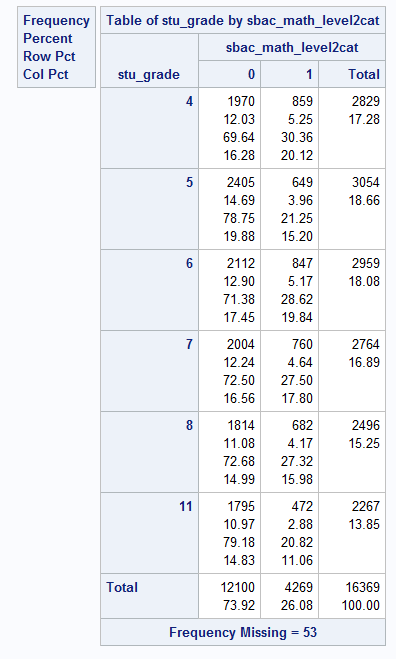


Frequencies for low SES and outcomes (ELA and Math) using SAS on cleaned\_combined\_all3 dataset



Frequencies for grades and outcomes (ELA and Math) using SAS on cleaned\_combined\_all3 dataset





When comparing with the R descriptive analysis and SAS analysis, there was no difference thus can indicate that the analysis between both programs are valid for descriptive analysis.

\*Model 1 with binary support aggregate data for English\*; \*references chosen

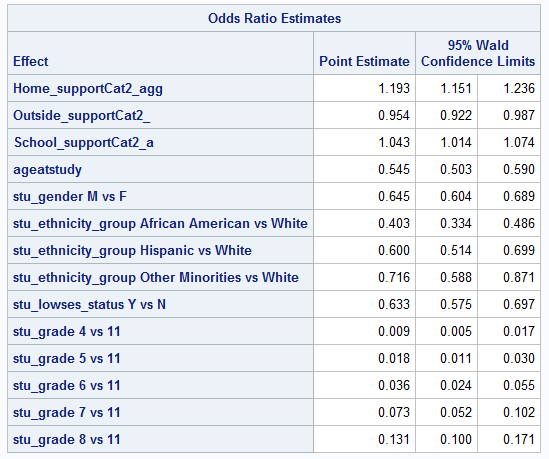
gender = female, females do better at tests according to descriptive data

race = whites, white do better at test according to research and descriptive data

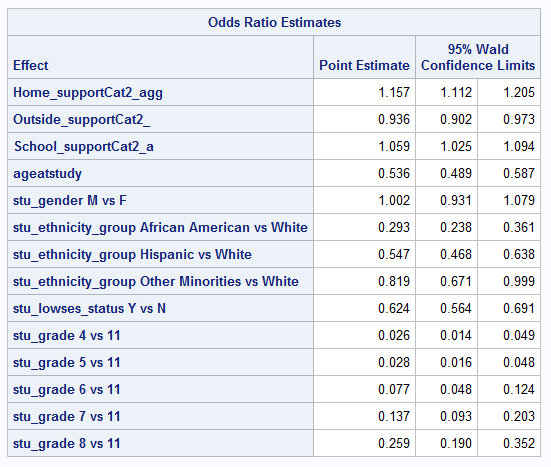
lowSES = No, individuals who are not in the low SES spectrum do better according to research and descriptive data

grade = 11, grade 11 had the strongest effect in OR when not used as referece, such this is chosen as the reference group\*

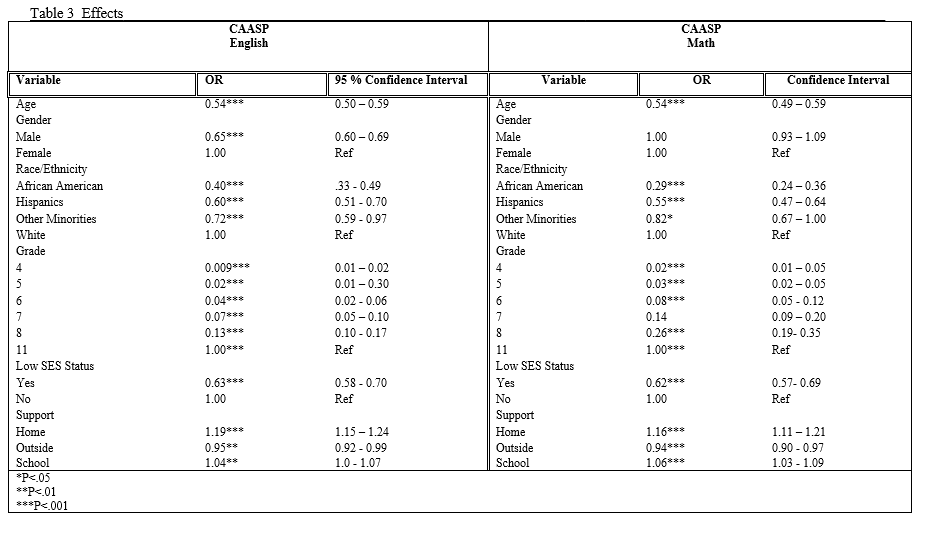
binary model with English as exposure



Binary Model with Math as exposure

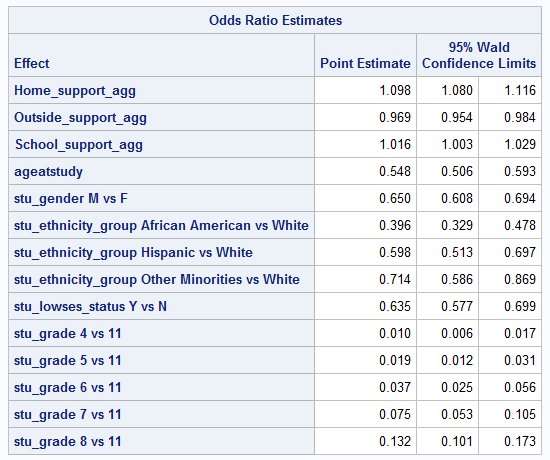


Fixed table with SAS analysis – error from last week was the coding

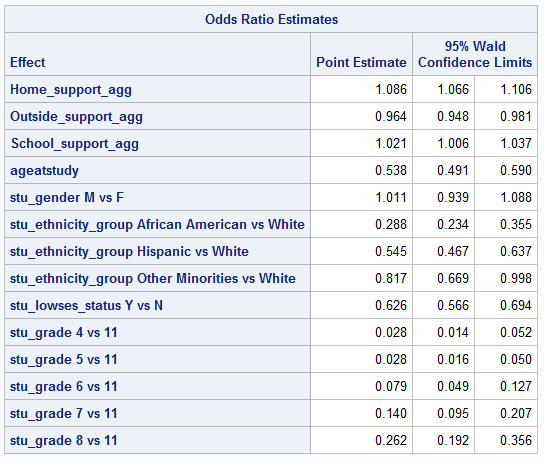


We see that in both CAASP math and English, we see that Home and School have a very significant effect towards test success.

Dr.Shavlik’s Model 1 continuous variables

English

Math



Now, Dr.Shavlik said to collapse the continuous 4 level factors into different categories. Refer to first page for that collapse.

\*collapsing continous variables

\*Home support\*;

**data** work.master2;

set work.master;

if **0** <= home\_support\_agg <= **6** then home\_supportcl = **1**;

if **7** <= home\_support\_agg <= **8** then home\_supportcl = **2**;

if **9** <= home\_support\_agg <= **10** then home\_supportcl = **3**;

if **11** <= home\_support\_agg <= **12** then home\_supportcl = **4**;

if **0** <= outside\_support\_agg <= **6** then outside\_supportcl = **1**;

if **7** <= outside\_support\_agg <= **8** then outside\_supportcl = **2**;

if **9** <= outside\_support\_agg <= **10** then outside\_supportcl = **3**;

if **11** <= outside\_support\_agg <= **12** then outside\_supportcl = **4**;

if **4** <= school\_support\_agg <= **8** then school\_supportcl = **1**;

if **9** <= school\_support\_agg <= **10** then school\_supportcl = **2**;

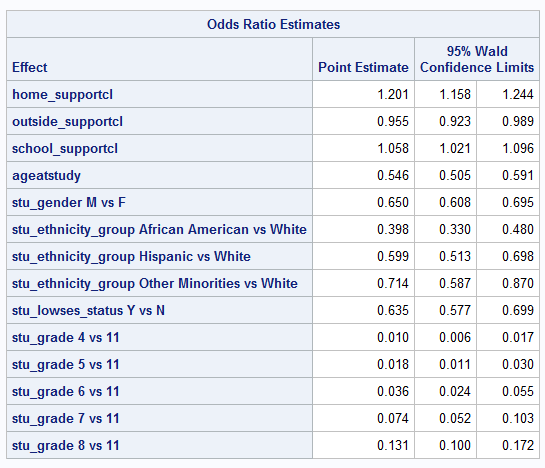
if **11** <= school\_support\_agg <= **12** then school\_supportcl = **3**;

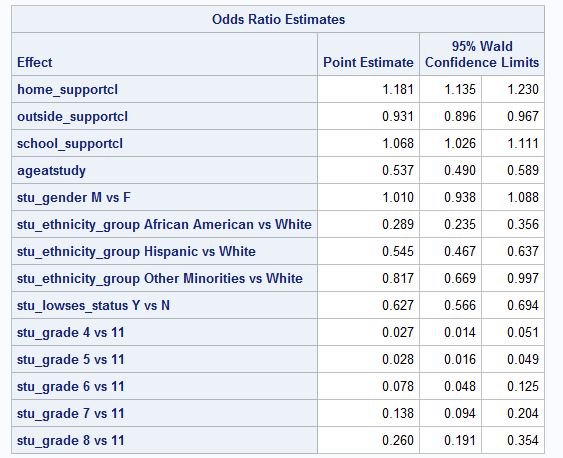
if **13** <= school\_support\_agg <= **14** then school\_supportcl = **4**;

if **15** <= school\_support\_agg <= **16** then school\_supportcl = **4**;

**run**;

English Model 2



Math model 2

MAPS –please refer to my SBCUSD powerpoint for maps

Maps have been improved.

There are 4 maps

* San Bernardino School District
* San Bernardino School District by Total Crime 2019 Index
* San Bernardino School District by Average Household income
* San Bernardino School District by Food Environment

4/13/2020

Multinomial analysis – look at levels 🡪 individual level 🡪 factor in different schools 🡪

SAS 🡪 individual level 🡪 where grade is 4 etc. 🡪 by itself 🡪 gender, low ses and support 🡪 where grade = 4 🡪 one of the subjects

Extra analysis - tried to look at some more 🡪 doing some grade level and age level

GIS

Overlay 🡪 Test scores by zipcode –

Crime rates

Households income

Fastfood

Food desert

Put in our test scores 🡪 Pass/Fail

Is it possible? 🡪 to get rid of that low income layer 🡪 that shows SES layer consist of our demographics 🡪 Figure out how to do that 🡪

Problem 🡪 how to show test performance on GIS map 🡪 ask Dr.Wiafe

PAPER 🡪 BIG PICTURE

Putting support on the GIS map

Morning Darelle,

Again, Great job on all your work this week.

Just a reminder for this week:

**Week of April 13th, 2020**

Regression

-       Just a reminder for this week, work on running the regressions again but per grade level; use "where grade=4;"

GIS

-       For GIS try to figure out how to create a layer for income based on our SES variables.

-       Also figure out how to overlay test scores and support on the map

-       See the below papers, I will also try to get a hold of it and read it too.

-       Kerski, J., Linn, S., & Gindele, R. (2005). Mapping standardised test scores with other variables using GIS. *International Research in Geographical & Environmental Education*, *14*(3), 231-236.

-       **Can we do different colored clusters for the test scores and support? see below link**

-       [https://doc.arcgis.com/en/arcgis-online/analyze/perform-analysis.htm](https://webmail.llu.edu/owa/redir.aspx?C=Uu-liJthH2agU2M8aWnq61yxLHwPAznBtt6bSSbn33BiCiXPwt_XCA..&URL=https%3a%2f%2fdoc.arcgis.com%2fen%2farcgis-online%2fanalyze%2fperform-analysis.htm)

4/15/202

Reasons for collapsing at the beginning 🡨 question for Nipher

Do the same for Math 🡪

4 models, Continuous and Cat

For the age 🡪 collapse age into 4 groups 🡪

put age continuous and age-squared

For both outcomes, evaluate for non-linearity

Check if the beta coefficients are linear

Need to be equally spaced

Once that done

Age collapse

9.8 to 11.99 -1

12 and 13 - 2

14 and 15 - 3

16 and 17 - 4

18 and above -5

5 categories for the two outcomes of english and math

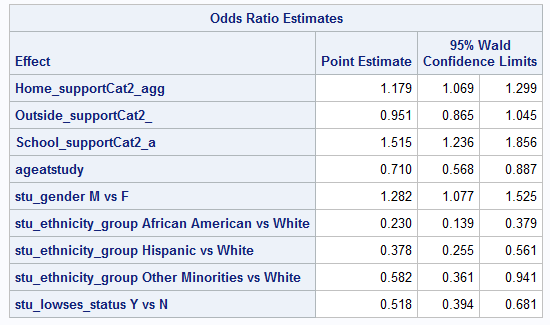
4/20/2020

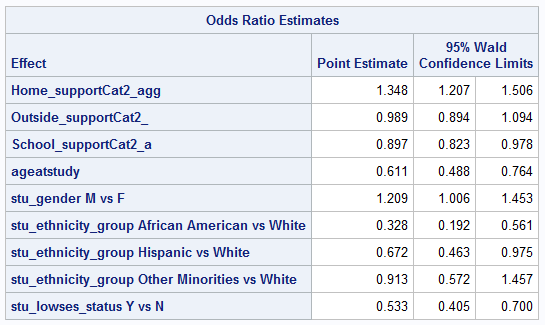
**Meeting with Dr.Shavlik**

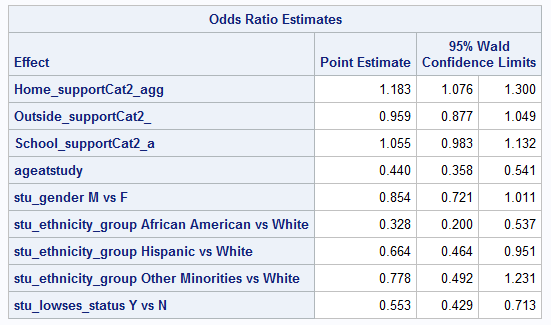
Finished the analysis part 🡪 can go on the next step 🡪 finishing literature review etc.

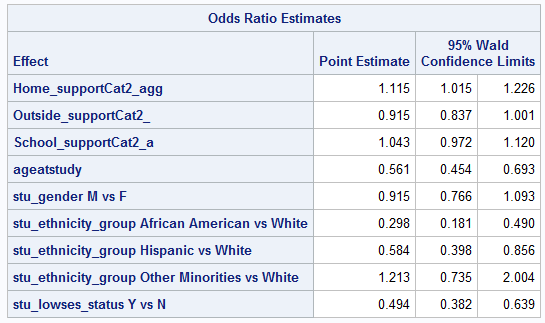
**SAS**

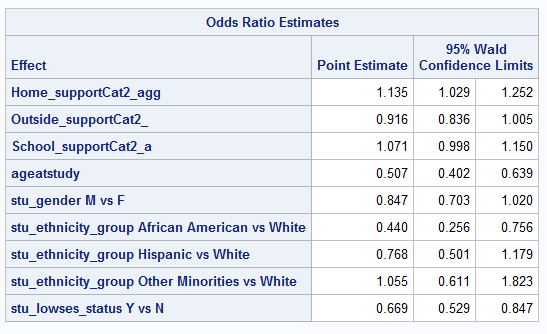
For each grade level

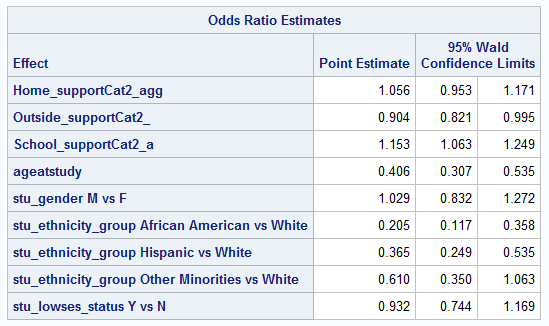
4

5

6

7

8

11

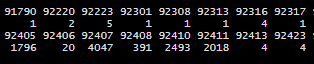
**GIS**

I created a new CVS file to work with the ArcGIS Pro program in order to use the counts of Low SES status and potentially test scores.

However,

In terms of GIS, there as something odd…some or many of the kids lived outside the SBCUSD district boundary. Using RStudio, I was able to pullup the zipcode where the participants answered from

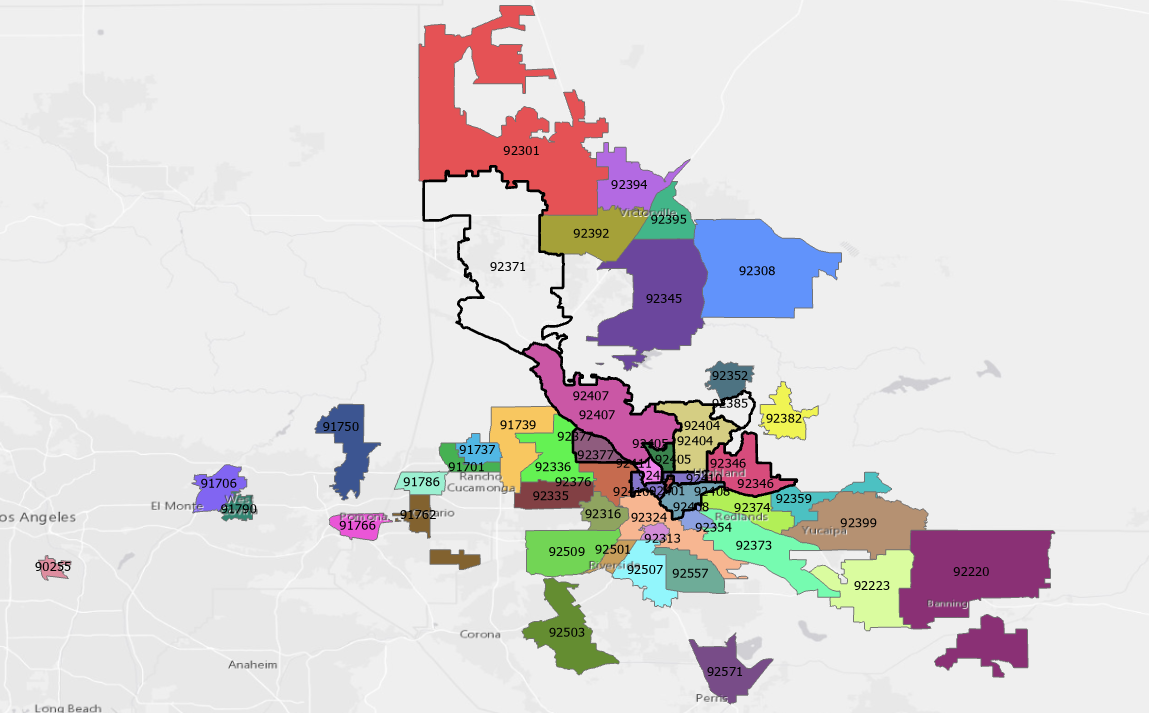








What I did next was select the zip codes on the GIS map accordingly



We see on the map, that many of the children live outside the zipcode boundary of SBCUSD. Next I need to put the lat and long of the GIS cvs file according to the zipcode.

One layer – Both living outside and inside

Green = passed , Red = failed 🡪 clusters

Zip codes were filled by parents (literature review 🡪 (kids outside sbcusd) 🡪 parents can be divorced

Math and English 🡪 map 🡪 percentage on the map

3 Maps

Math

English

SES

**5/18/2020**

Why is age group – categorized –

Pick one tab

ask Dr.Shavlik