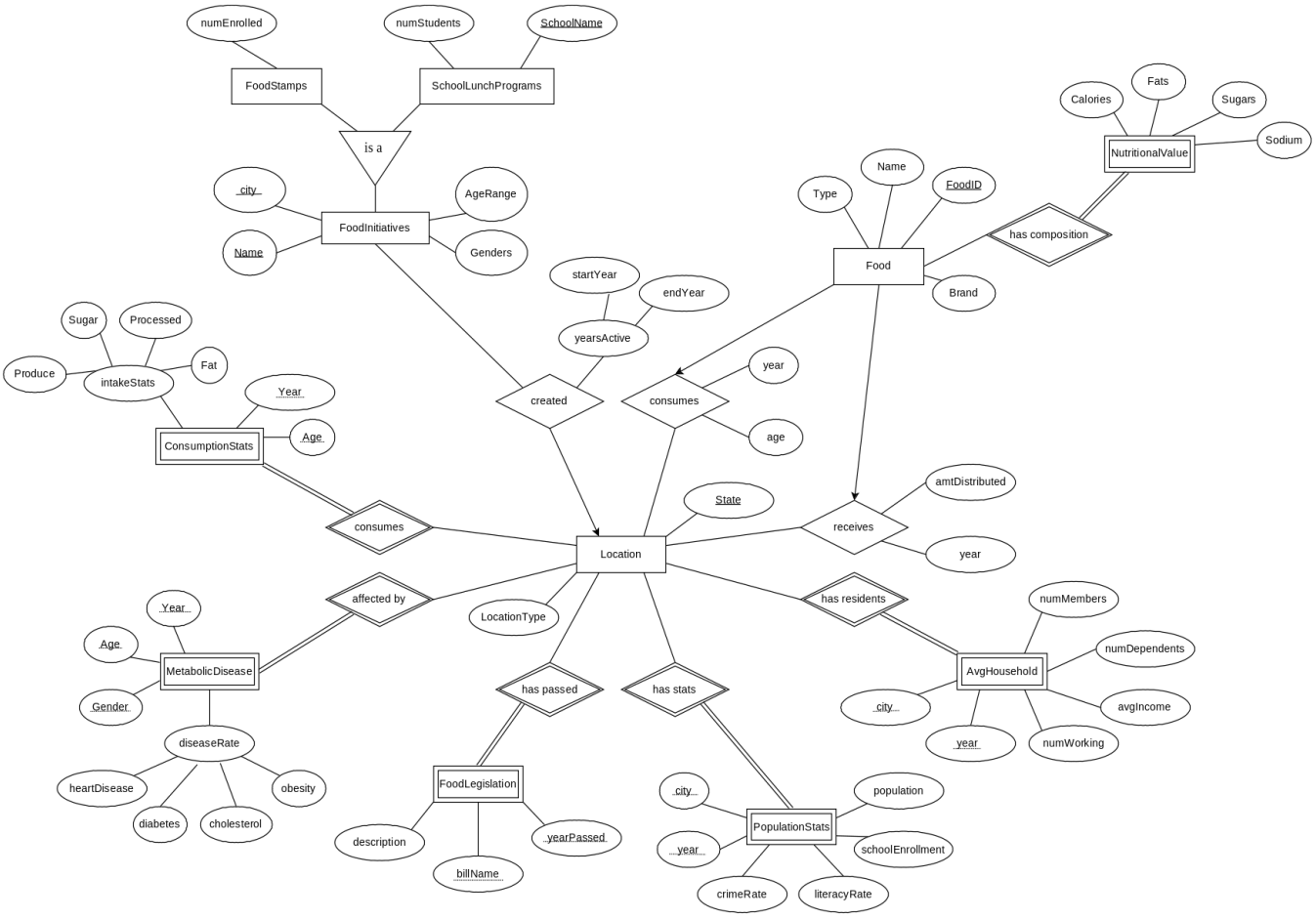
# Project Phase B: ER Diagram and Relation Schemas

**Partners: Louie DiBernardo (ldibern1), Sanddhya Jayabalan (sjayaba1)**

## Part 0: Updates

One of the comments given to us about our data sources was to try and find similar CA research studies for the entire US. We looked through several states' databases across the county and none had the same data sources as CA's. Currently our plan is to perform nationwide and state-specific analysis based on the type of question we are trying to answer as well as the data available. For example, if we want to answer a specific questions about school lunch programs at a state level, we will use the CA data. However, we will try to find nationwide data to answer as many of our questions as possible and adapt questions accordingly based on data availability.

## Part 1: E-R Diagram



## Part 2: Draft Schemas

Location(*State, locationType*)
FoodLegislation(*State, billName, yearPassed, description*)
PopulationStats(*State, City, Year, CrimeRate, LiteracyRate, SchoolEnrollment*)
AvgHousehold(*State, City, Year, NumMembers, NumDependents, NumWorking, AvgIncome*)
ConsumptionStats(*State, Age, Year, SugarIntake, FatIntake, ProcessedIntake, ProduceIntake*)
MetabolicDisease(*State, Age, Year, Gender, diabetes, obesity, cholesterol, heartDisease*)
FoodInitiatives(*State, City, Name, AgeRange, Genders*)
FoodStamps(*State, City, Name, numEnrolled*)
SchoolLunchPrograms(*State, City, Name, schoolName, numStudents*)
FoodInitiativesCreated(*City, State, Name, startYear, endYear*)
Food(*FoodID, Name, Brand, Type*)
NutritionalValue(*FoodID, calories, sugars, fats, sodium*)
FoodConsumption(*State, Year, Age, FoodID, amtConsumed*)
FoodDistribution(*FoodID, State, Year, amtDistributed*)

We made some simplifications with the composite attributes when designing the schema where rather than including an additional relation for all the sub-components of the composite attributes, we chose to include them as attributes in the main relation (ex: for MetabolicDisease). In addition, rather than having a separate relation to include the year a certain bill was passed in FoodLegislation, we chose to combine this value as an attribute in the FoodLegislation relation itself.

## Part 3: Finalized Schemas

- Unique identifiers: We played around with the idea of introducing unique identifiers for representing location and food items. We decided to keep the foodID value to be assigned to each individual food item because we felt this would help remove complications of requiring multiple attributes to represent a single food item, for example if multiple brands made a producct with the same name, or if the same brand made a product with the same name over different years. We are struggling with how to best represent locations in our database. Some of the datasets we found represent data at the state level, while others break the data down to the county level or further down to the city levels. We want to include data with different levels of location breakdown in our database so we are unsure what would be the best way without introducing too many null values. We originally planned on using a locationID attribute in the Location entity set but ended up removing it and replacing it with more weak entity sets; however, we are still little unsure of this design. This is one area we would really like feedback on as we move into the next steps of the project.

- Functional Dependencies: All the functional dependencies currently present in our schemas contain primary/candidate keys on the LHS going to non-primary attributes in the RHS suggesting that our schema are in 3NF form. Despite the normalization, we would like to consider making some changes to our schema to make it more simple especically in how we represent location. We believe our current schema could potentially represent all desired location values but may be doing so in a complicated way. We would like to

continue thinking of some ways to simplify location representation potentially by using a locationID but we need to make sure we preserve the ability to have different data that are broken down to different leels of location (i.e. state-level, county-level, city-level, etc.).

- Other constraints:
  - Foreign key constraint where value of state attribute in any relation other than Location must be present in Location relation (later if we end up changing to using locationID, then this foreign key constraint will be replaced by a similar one requiring locationID present in any relation other than Location to be present in Location)
  - All rate/amount values must be non-negative in any relation (ex: amtDistributed in FoodDistribution, amtConsumed in FoodConsumption, calories in NutritionalValue, etc.)
- Final schema: same as proposed in part 2 (would like some feedback on how to represent location)

### Data Sources

National School Lunch Program: State Implementation Progress, School Year 2012–2013: Report to Congress
Agricultural Productivity in the U.S.
CDC Nutrition, Physical Activity, and Obesity - Legislation
SNAP Income Limits
Supplemental Nutrition Assistance Program Participation and Cost Data
Supplemental Nutrition Assistance Program (SNAP) Data System
Let's Move Salad Bars to Schools Final
Rates and Trends in Hypertension-related Cardiovascular Disease Mortality Among US Adults (35+) by County, Age Group, Race/Ethnicity, and Sex – 2000-2019
Women, Infants and Children (WIC) Program: Penetration Rates 2009-2012
Heart Disease Mortality Data Among US Adults (35+) by State/Territory and County – 2017-2019
Obesity in California, 2012 and 2013
Fruit and Vegetable Consumption in California Residents, 2012/2013
Sugar-Sweetened Beverage Consumption in California Residents
Nutrition, Physical Activity, and Obesity - Women, Infant, and Child
Public School Enrollment
Food Affordability (CA)
California Adults Who Met Physical Activity Guidelines for Americans, 2013