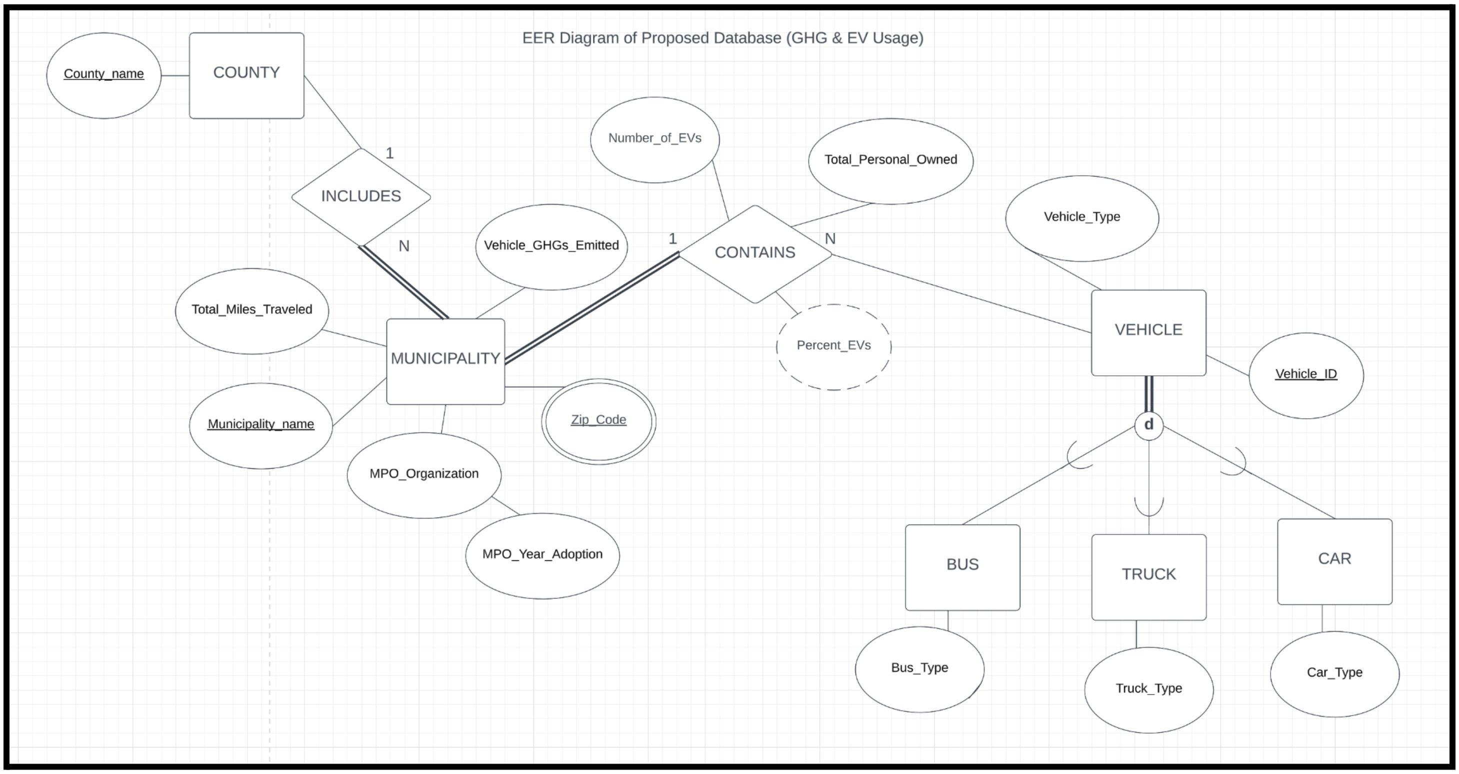
Madison Bavosa, EJ Gasataya, Kristen Libonati, Jessica Malek, Emmanuel Pasteur

Phase III - The Database Model Explained

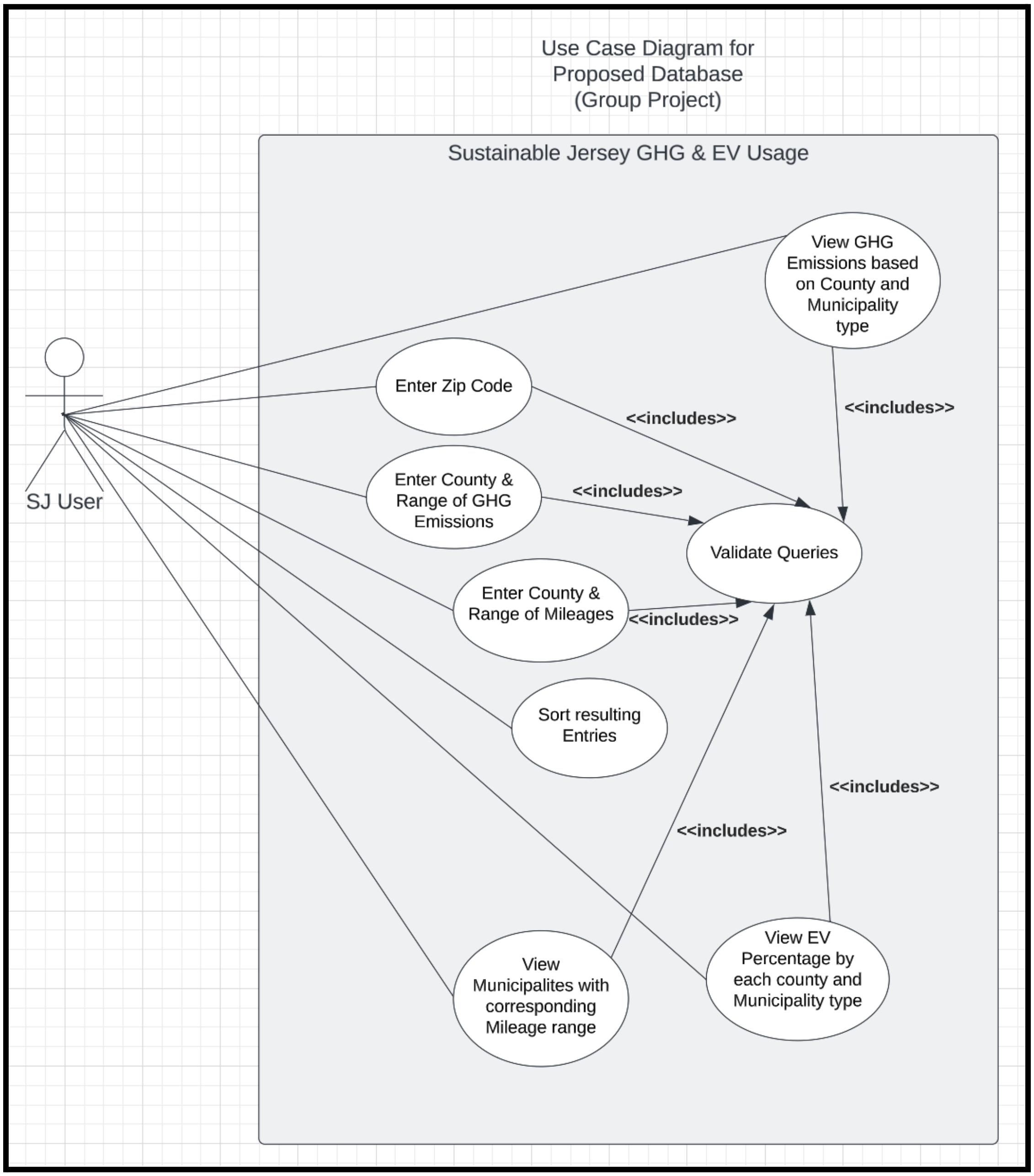
Professor Michels & Professor DeGood | MGT/BUS-385 & CSC 315 | 10 March 2023

ER (EER) Diagram of Database



(Use Case Diagram on next page)

UML Use Case Diagram of Proposed User Interaction



<u>Summary of the Details of Our System's Users (Actors), Their Interactions with Our System, and the Goals Our System will Help Those Actors Achieve</u>

The user interface will consist of several input forms that the user will enter into the system as a query (request). More specifically, the user will have to enter a county in NJ (and/or zip code), which will output a list of municipalities within the county that shows information regarding GHG and EV usage. The user will also have the ability to sort resulting entries from the query by ascending or descending order. For example, this can be useful by identifying which municipalities have the *highest number of miles traveled but have a low population of EVs, alongside the GHG emissions within the municipality.* Validations of queries (which will be done internally within the database) will also need to be taken into account; the user interface can have a message saying that the user will need to enter appropriate queries- for example, a ZIP code cannot be negative, have more than 5 digits, etc. We also need to ensure the range of values for GHG and miles traveled.

Scope of Our System

The scope of the system will heavily depend on which datasets we are going to use. As described in our proposal, we are going to be focusing on New Jersey datasets, collected from 2020-2022. Thus, we will have to make assumptions that when information is being presented in the user interface, we have to remind the user that the result of queries will be on datasets from NJ. If we have to take into account GHG emissions and EV percentages *nationwide*, then the scope of the system will be larger. We also decided to ignore motorcycles and motorhomes since they contribute very little to the total vehicle miles traveled and GHG emissions.

(Textual Use Case Descriptions are at the end of this document)

Narrative:

How Relational Databases Work and Why They are Valuable

A relational database will store a collection of information that can be represented as “relations”, and the relationships they participate in. Relations are based on the entities drawn from ER/EER diagrams, representing things we see as everyday objects. Some of the reasons why relational databases can be beneficial are because they can provide an in-depth view of how commonalities across different datasets can be drawn out. Also, the type of actions you can do with relational databases is very flexible yet powerful. The ability to easily update, delete, or add tables and/or relationships allows the database to adapt to datasets that are very complex. Data integrity is another valuable aspect of a relational database; it is important that there is consistency among each instance of the entity in terms of the *domain* constraints held in each one. For example, there needs to be a mechanism in the database where no user can query an invalid zip code (does not fit within the attribute domain), or be able to modify it to an unreasonable amount.

Elements of Our Diagram and What They Reveal About Our Database Model

Our ER/EER diagram includes five entities. These include “COUNTY”, “MUNICIPALITY”, “VEHICLE”, “BUS”, “TRUCK”, and “CAR”. “COUNTY” represents any one of the 21 counties in New Jersey. Each county is identified by its name. “MUNICIPALITY” represents the different townships in New Jersey. Each municipality is identified by its name and corresponding zip codes. This entity also includes information such as the total vehicle miles traveled and the municipal planning organization. The “VEHICLE” entity will represent a vehicle. Since we do not have details such as license plate numbers from our selected data sets, we will use a *surrogate key* to identify different vehicles. Additionally, the “VEHICLE” entity will have three sub-entities, according to the “Vehicle_Type” attribute. (we will use attribute-defined specialization). These are “BUS”, “TRUCK”, and “CAR”. We assume that a vehicle can only be either a bus, truck, or car (but not both). Each of these entities will have an attribute that specifies the type of bus, truck, or car the item is. For example, the domain for “Truck_type” is combination long-haul, combination short-haul, single unit long-haul, single unit short-haul, refuse, passenger, and light commercial.

The relationship type “CONTAINS” will have attributes describing the participation between each municipality and the number of vehicles. The number of EVs will be drawn out from this relationship as well as the number of total vehicles owned. The “Percent_EVs” attribute is going to be used as a derived attribute since it can be calculated based on the *ratio* between the number of EVs and the total number of personal vehicles owned in each municipality. We also suspect that the zip code attribute in the “MUNICIPALITY” entity will be the primary key *as well as being m*ultivalued. This is because when we were looking at our datasets, we noticed that some municipalities contain different zip codes at the same time.

Reasoning Behind Our Database Given the Goals for the Sustainability Project

The design of our database can be leveraged to deliver our proposed sustainability goals of recognizing which municipalities have the highest number of electric vehicles compared to municipalities that have the lowest number of electric vehicles, comparing that data to total vehicle miles traveled and in turn analyzing the ratio of vehicle emissions to vehicle type (EV vs Non-EV). The database will enable an effective way for users to identify the vehicle miles traveled for each municipality and the proportion of electric and gas-powered vehicles per municipality, as well as the total emissions associated with them respectively.

Textual Descriptions based on EER Diagram:

Use Case 1: Specifying Range of GHG Emissions: the user can enter a range of GHG emissions as a query, and return the counties that fall within that range.

1. The system will prompt the user to enter which county they want to analyze.
2. The User selects an available county (the domain is only in NJ)
3. Within the county selected, the user will then enter a range of values that represent GHG emissions.
4. The user must enter the range of GHG emissions in metric tons.
5. The system validates the range of values from the user.
    a. The range is invalid (i.e User enters a negative value)
    b. The system tells the user that the range is invalid
    c. The use case will go back to step 4 until the user enters a valid range
6. The system will present a list of municipalities (with their corresponding zip codes) with the selected county that falls in the GHG emission range.

Use Case 2: Specifying Zip Code to View EV Percentage: A user will be able to select from the municipality list as mentioned in Use Case 1. Then, the system will allow the user to view the number of electric vehicles owned in the selected town. The user will also be able to see the percentage of owned vehicles in the township that are electric.

1. The system prompts the user to enter a zip code.
2. The user enters a valid zip code.
    a. The zip code is invalid (i.e. User's input is less than 5 digits)
    b. The system presents an error message to the user
    c. The user case will go back to step 1
3. The system displays the specified town and asks the user to verify it.
4. The user chooses to verify or deny the town.
    a. The user denies the town
    b. The user case will go back to step 1
5. The system displays the number of electric vehicles in the verified town as well as the percentage of electric vehicles in comparison to the total number of vehicles owned in the township.
6. The system will present a "status" message when the percentage of EV usage is pretty low relative to the EV sample size in each municipality.

<u>Use Case 3: Specifying Mile Ranges:</u> This functionality is going to be similar to how we enter data on GHG emissions data. The user will have the option to also enter the mileage range they want, which will return the list of municipalities (from a given county) that fall within that range.

1. The system will prompt the user to enter which county they want to analyze.
2. The User selects an available county (the domain is only in NJ)
3. Within the county selected, the user will then enter a range of values that represent the mileage traveled.
4. If the distance is given in different units (i.e meters), then they must convert this into miles.
5. The system validates the range of mileage from the user.

    a. The range is invalid (i.e User enters a negative value or incorrect units)
    b. The system tells the user that the range is invalid
    c. The use case will go back to step 4 until the user enters a valid range
6. The system will present a list of municipalities (with their corresponding zip codes) with the selected county that falls within the specified mileage range.

<u>Works Cited</u>

"Car Emissions and Global Warming." *Union of Concerned Scientists*, 2014, https://www.ucsusa.org/resources/car-emissions-global-warming.

EPA, Environmental Protection Agency, https://www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-vehicle#:~:text=typical%20passenger%20vehicle%3F-,A%20typical%20passenger%20vehicle%20emits%20about%204.6%20metric%20tons%20of,8%2C887%20grams%20of%20CO2.

Lindsey, David Herring, AND Rebecca. "How Will Global Warming Harm Human Health and Well-Being?" *NOAA Climate.gov*, 29 Oct. 2020, https://www.climate.gov/news-features/climate-qa/how-will-global-warming-harm-human-health-and-well-being#:~:text=The%20most%20vulnerable%20people%E2%80%94children,on%20a%20sweltering%20Wednesday%2C%20Aug.