

# **Mid-Semester Collaborative Project Report (Group 4)**

Group Members: Jason Kantner, Gordon Petry, Michael Williams, Casey Lishko, Yanaja Joyner, Len La Rocca

## **Stage I: Team Formation (Group 4)**

- Group information is up to date on the Canvas group wiki page

---

## **Stage II: Project Proposal and Specifications (Group 4)**

Group Members: Jason Kantner, Gordon Petry, Michael Williams, Casey Lishko, Yanaja Joyner, Len La Rocca

[Link to GitHub Wiki](#)

### **Problem Statement**

For this project, the main problem that we want to address and find a solution for is the lack of easily accessible/understandable sustainability data on the CivicStory website. Currently, the CivicStory website is a hub for relevant news articles relating to sustainability, however direct information on sustainability efforts based on geographic location is not accessible to the common user.

### **Objective of the Module**

The objective of this module is to create an addition to the CivicStory website that will allow normal users to access detailed data on sustainability in their area. Another objective which we hope to accomplish is to ensure that this data is easily understandable to the common user and it is tailored to what information they want to know. It is our hope that this tool will keep the public well informed about sustainability in their own area.

### **Desired End Product / Part Developed for This Class**

In general, the product we plan to design can be used by many different websites and is not only limited to CivicStory. The module that will be designed will be a page where the user can choose from many different variables relating to data in a connected database. Once the user specifies the data they want to see, they then get to choose a graph layout for the data to be visualized as (Ex. pie chart, bar graph, line graph, etc). After all of the selections are complete, queries will pull data from the connecting database in PostgreSQL and the desired data will be

visualized onto the screen of the user. For this class, the data in the database will be relating to sustainability in the area. This makes the planned module a great tool to be utilized on the CivicStory website because it strongly relates to the purpose of the website and the goals it wants to accomplish.

### **Importance of this Module and How it Addresses the Problem**

Right now CivicStory is mostly a stream of news articles relating to sustainability efforts. Although informative, these do not offer the user a look at raw data that relates to the issue. Allowing the user to search for specific data gives them a true understanding of the issues revolving around sustainability and what can be done to improve it in the future. Also, they can compare and contrast data from different times or places to get the “whole picture” of where sustainability is, where it comes from, and where it needs to go.

### **Problem Domain Research Plan and Data Obtainment**

To research the problem domain, we will look over the CivicStory website to see what type of data would be relevant to put in the database connected to the website. Not every piece of sustainability data we find should be put in the database. For example, we need to have a limited geographic location so it remains relevant to those that will access the CivicStory website. Once we research the problem domain in more depth, it will then be much easier to figure out what kind of data we will implement into our database. We will obtain this data from the Sustainable Jersey website. This website has a large amount of data regarding sustainability in New Jersey divided by counties. We will collect this data and store it in our PostgreSQL database.

### **Similar Systems and How our System Differs**

Search bars and user interfaces that allow for specific search terms do exist in many forms (Ex. PAWS class search tool). However, the addition that we plan to make ours stand out from others in that our search tool creates a data visualization based on the information provided by the user on what data they want to view.

### **Other Possible Applications for this System**

The search system has wide applications on many databases. It is a robust search system that visualizes the data, which is useful in any database. Possibly, instead of typing in the search criteria, the user can point out on a map where they want to search, or use a slider to determine what time period they want to look at. However, the data visualization tool will always be used after the search.

### **Performance**

The system we plan to design will ensure efficiency between the many programs that are being utilized. Our goal is to reduce lag time as much as possible between the Database

Management System and the web interface. This way, when the user prompts to create a data visualization, the task will be completed in a reasonable amount of time. Performance and efficiency should always be a top priority when creating a new web tool to use. This will be accomplished by taking time and planning out our code before going into the programs right away. We will also continue to test our system throughout the development process to see what areas we need to improve.

### **Security**

Regarding the security of our system, we don't plan to allow users to directly alter the data within the PostgreSQL database. This access will be limited to only administrators, who are allowed to do so. For the users, they will only be able to select the parameters in which they want to query the data and view the visualization that it provides. So, the users are only limited to viewing the web interface side of this module.

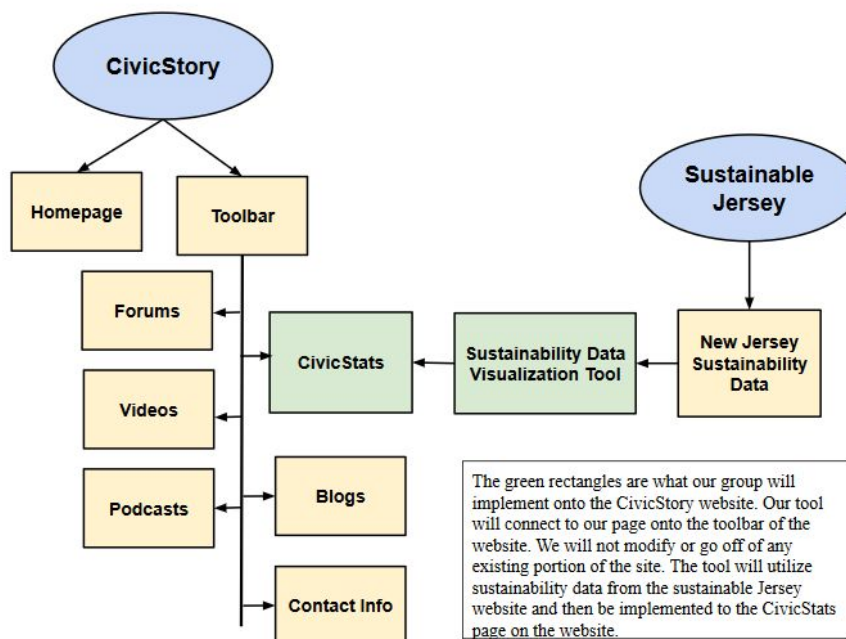
### **Backup and Recovery**

Backup and recovery are an important part when implementing a tool that utilizes data. For our project, we plan to have a copy of the database that will be connected to the web interface. This means that if anything happens to the database connected to the website, we will have a copy of the data just in case it needs to be utilized. This copy database will also be within PostgreSQL which makes it readily available to replace the original database if need be.

### **Technology and Database Concepts to Learn**

For this project, there are a few database concepts and technologies that our group will need to learn to be able to accomplish our objective. First, our team will need to learn how to utilize the Database Management System PostgreSQL. It is within PostgreSQL where our database containing the sustainability data will be located. We will also utilize HTML to create the web interface side of this project. This will create the frontend of the project where the user will view the data visualization tool. To implement this data visualization tool, we will also learn how to use data visualization software. Data visualization software is abundant and is also vital to properly creating our module. Finally, Python will be learned and utilized to bridge the gap between the web interface and the Database Management System. Python is very important to learn because without it, we would not be able to make the connection between the front and back end portions of this module.

## Diagrammatic Representation



## Quad Chart



### Sustainability Data Visualization Tool

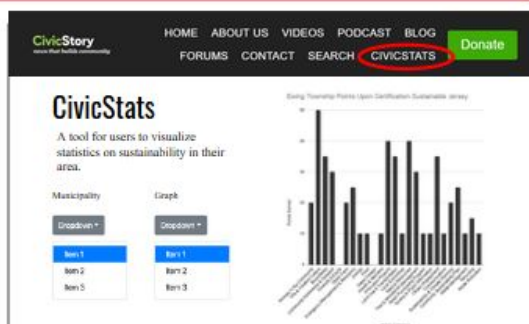
Jason Kantner, Gordon Petry, Michael Williams, Casey Lishko, Yanaja Joyner, Len La Rocca

#### Objective

Develop a system that converts sustainability data into easily understood visualizations.

CivicStory does not currently have resources on sustainability data and this will allow users to choose what data they want to see and in what form.

We hope to inform common users about sustainability and how it impacts their daily lives.



#### Approach

Modify the existing CivicStory website by

1. Creating data visualization software that will allow users to see specific data represented in a graph, pie chart, etc
2. Creating drop down menus that allow the user to select what type of data is visualized and in what form (graph type) it is shown.
3. Utilize tools such as PostgreSQL and Python to streamline the software so it can run efficiently and in a timely manner.

#### Key Milestones

- Proposal & Pitch ----- 2/12/20
- Specifications ----- 2/24/20
- Model ----- 3/09/20
- Design ----- 3/26/20
- Tables and Queries ----- 4/09/20 in PostgreSQL
- Implementation ----- 4/27/20
- Demos & Presentations ----- 5/04/20
- Final Report ----- 5/04/20

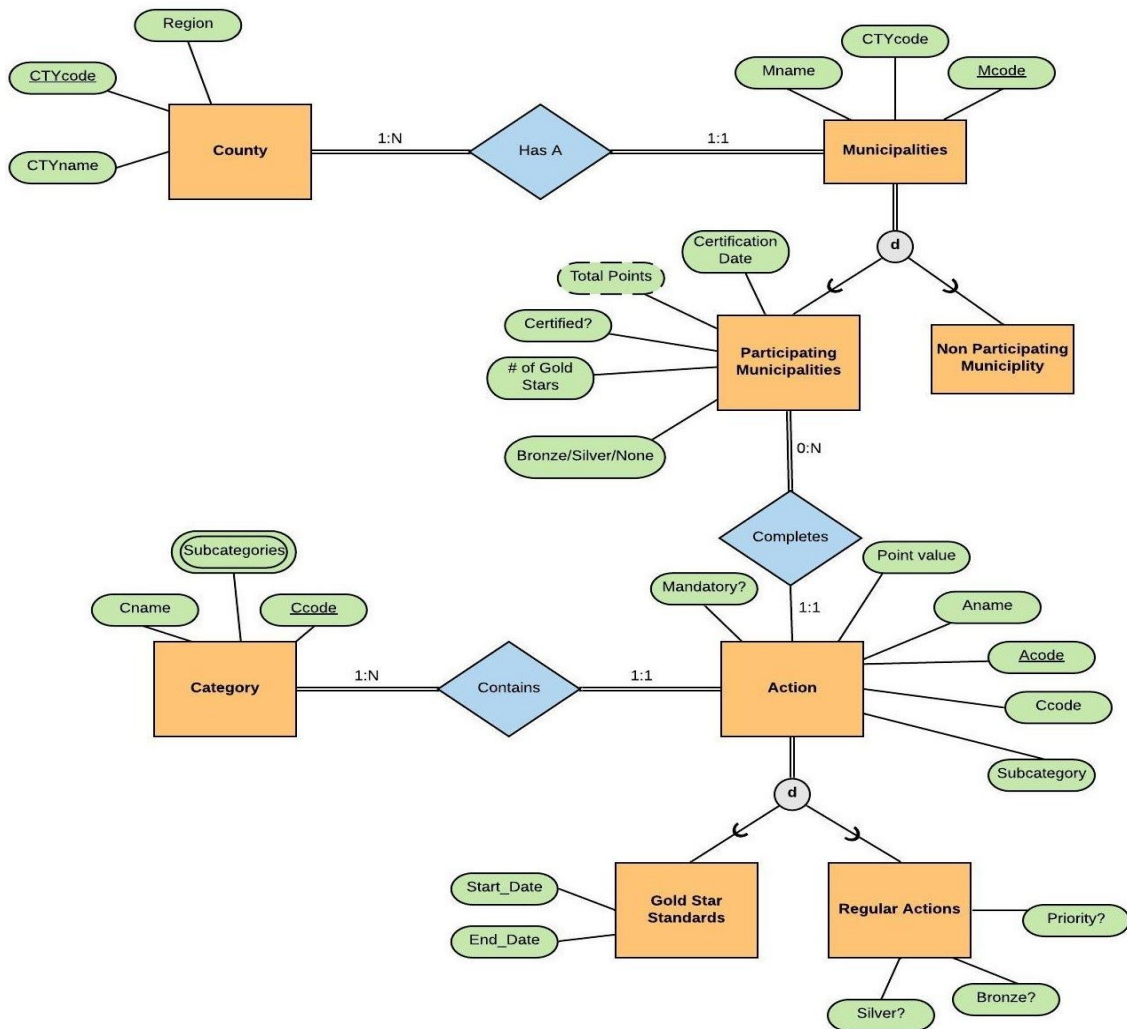
---

## Stage III: Data Model (Group 4)

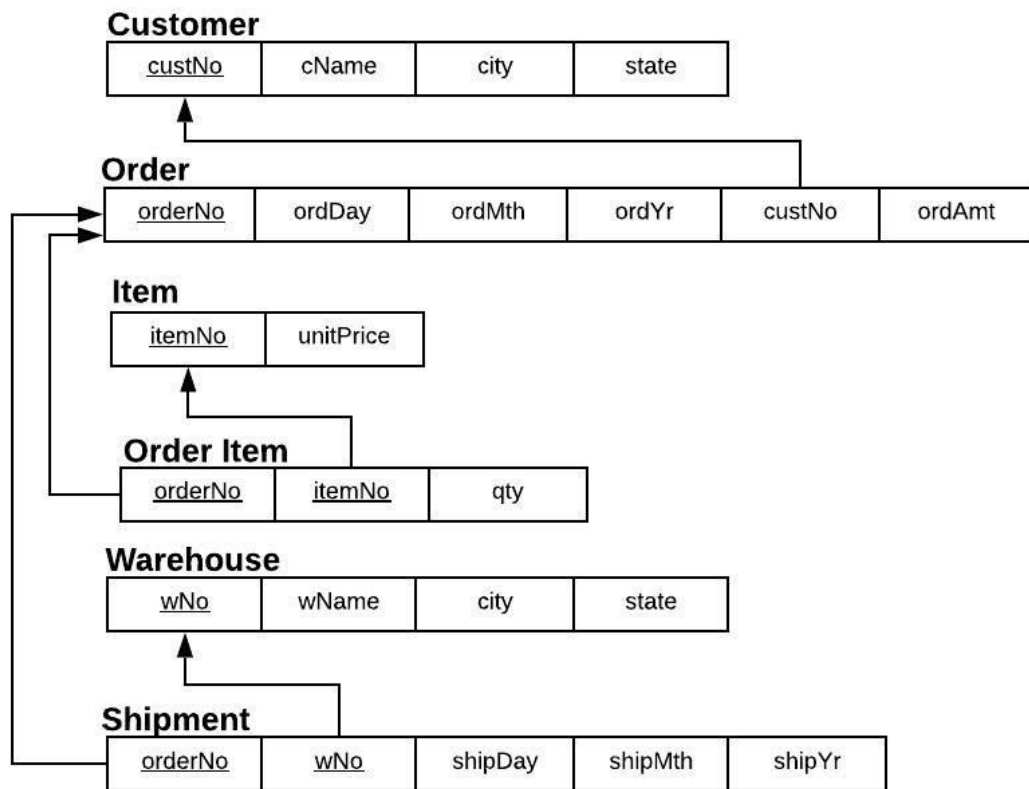
Group Members: Jason Kantner, Gordon Petry, Michael Williams, Casey Lishko, Yanaja Joyner, Len La Rocca

[Link to LucidChart](#)

### *ER Diagram*



### *Relational Schema*



## ***Additional Information***

- Initial Database Size (approximate number of records)
  - Our initial database hold **750** records
    - Counties: 21
    - Municipalities: 565
      - Participating Municipalities: 455
      - Non Participating Municipalities: 110
    - Actions: 146
      - Gold Star Standards: 2
      - Regular Actions: 144
    - Categories: 18
- Types and average number of searches
  - Our initial estimate of queries that will be used in a typical user session is around **10 queries**

- We got this rough estimate because we are planning to have multiple drop-down menus where users can choose many variables to filter the data to their liking. The menus we currently have are below...
  - Graph Type
  - Municipality / Location
  - Sustainability Point Amount
  - Gold Star?
  - Silver Star?
  - Bronze Star?
  - Certifications
  - Certification Date Range
  - Actions Taken
  - Categories of Entries
- *Note: We are likely going to implement a way for users to generate up to 2 charts at once so they can compare data. This would bring the potential query usage per session up to around 20 queries*

---

## ***Inceptions***

### **Inception: Team Identification**

<b><u>Project Title: CivicStats</u></b>			
<b><u>Group Name: CivicStatisticians</u></b>			
<b><u>Name</u></b>	<b><u>TCNJ Email</u></b>	<b><u>GitHub ID</u></b>	<b><u>Phone #</u></b>
Gordan Petry	petryg2@tcnj.edu	gpetry37	(732) 966-3000
Michael Williams	willim48@tcnj.edu	mikeacyle	(732) 678-7821
Jason Kantner	kantnej2@tcnj.edu	jmkanter9	(856)701-3027
Casey Lishko	lishkom1@tcnj.edu	WaterCasey	(201) 468-1881
Yanaja Joyner	Joynery1@tcnj.edu	YanajaJ	(848) 241-4573
Len La Rocca	Laroccl2@tcnj.edu	...	(732) 804-0257

### **Inception: Objectives and Overview**

When our group was first formed and given this project, we were not entirely sure what we wanted to do. The first meeting along with the project requirements provided us with a lot of information at once. In addition to this, we were also given a lot of freedom to be creative with what our project would look like. This freedom and information initially made it difficult for us to come up with an idea that we wanted to approach. However, after much discussion as well as further research into this topic, we found a design that we believe will be a wonderful resource for the CivicStory website. Based on the information provided and what CivicStory is looking to add, we decided as a group to create a data visualization tool that allows users to easily access information regarding sustainability in the New Jersey area.

The reason why the data visualization tool is what we are pursuing originates at the start of this process. Our initial goal of this project was to listen to the recommendations given by the CivicStory representatives and to center our database around what they wanted to add to the website. During the first meeting, Kevin from CivicStory talked about how they wanted to have more data on the site and make it accessible to the common user. Currently, CivicStory is mostly a hub for news articles related to sustainability. Kevin also mentioned that there was a data collection site around TCNJ that collected information regarding sustainability in the New Jersey area. Once we heard about this data hub, we knew that this was definitely something we wanted to include in our project. Since we wanted to make this data easily accessible / understandable by the common user we came up with the idea of a data visualization tool. We aspire to make a database of sustainability data which has an easily understandable user interface to allow users to access data regarding sustainability that they would otherwise never find.

Along with wanting to create this tool to improve upon the CivicStory website, there is also a lot that we hope to learn and get out of this project. In the realm of computer science, we are looking forward to creating our own database and connecting it to a user interface. This project will help us better understand how databases are imperative when it comes to the creation of websites. This project will also help us better understand how to outline and implement a database from the ground up. This will most certainly be helpful both for our education and our future careers. We hope to learn to apply PostgreSQL to more than just the educational setting than what we are learning in class. We aim to hook up a database to code written in Python and



explore methods of displaying this on a website whether it be on CivicStory or as a standalone site. We want to do more research into tools that will help us do this, whether it be Python's Django, HTML5, or some other method. Additionally, we are looking into using Python's BeautifulSoup as a web scraping tool in order to gather the information we need for our database more efficiently. The best way to learn a new coding language is to implement it and that is exactly what we hope to get out of this project. Also, we are looking forward to working with the journalism class so we can connect these computer science topics to the real world. After all, the realm of computer science is all about improving systems to better assist the needs of people in the real world. Having the journalism majors as well as working with the CivicStory individuals will help us bridge the gap between the coding side of computer science and the real world implications it holds.

Originally, we wanted to simply make a system where a user could look for sustainability data by filtering it to their liking and the page would return a spreadsheet/table with the data they requested. This was our first approach when coming up with a system that allows users to easily access data regarding sustainability in their area. While this was a good start, we believed it was too basic and normal users may not fully understand the data they are looking at simply from the table. This led to the concept of making a visualization tool rather than simply printing out tables. We like this approach much more because we believe this is a great way for users to access the data they desire as well as fully understand the data they are looking at. Data visualization tools are a large part of the data analysis field. It helps both professionals and non-professionals better understand the data they are looking at thanks to the visualization rather than having to understand it through a lot of numbers on a large spreadsheet. Even though we moved away from our original concept, it was definitely an inspiration for our current model and without it we may not have come up with this idea to begin with.

Not only are we excited to further develop this project and better understand how to effectively implement databases, we are also glad to be working with the journalism majors. Even though they cannot help very much with the coding side of this project, they are an invaluable resource when it comes to planning and connecting this to the community. Developing a database is great, but it is not truly noteworthy unless it has an impact in the real

world. We believe that our CivicStats tool will make a positive impact on the CivicStory website. We are looking forward to seeing the effects our deliverable has and potential additions that could be implemented to make it even more impressive. We hope to take a lot out of this project that we can use in the future when it comes to implementing databases and solving problems in the real world.