Data Science Program

Capstone Report – Fall 2 024

Carbon Capture Technology

Emily Fugh

supervised by

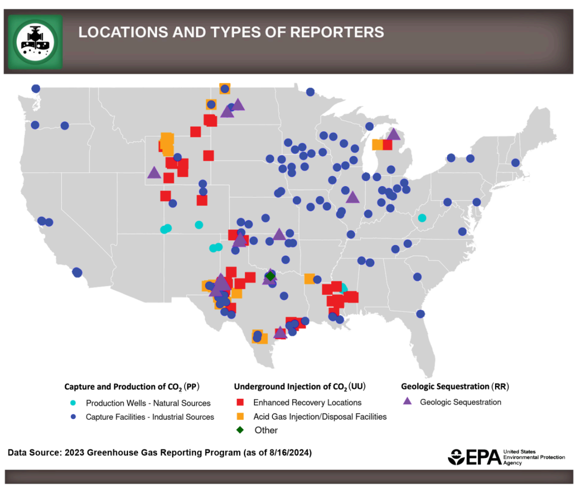
Sushovan Majhi

Introduction

What is Carbon Capture, Usage, and Storage?

Carbon Capture, Usage, and Storage (colloquially known as CCUS) is defined by the London School of Economics and Political Science as “a suite of technologies that enable the mitigation of carbon dioxide (CO_2) emissions from large point sources such as power plants, refineries and other industrial facilities, or the removal of existing CO_2 from the atmosphere” (“What is Carbon Capture?”). For the purpose of this project, we are only interested in the technology that allows CO_2 to be **stored** underground, whether that be for solely keeping it out of our atmosphere or for other purposes. This technology has gained a lot more attention in recent years as efforts to control climate change and reverse our damage on the planet intensify.

In the United States, organizations that are interested in carbon storage report to the Environmental Protection Agency (EPA)’s Greenhouse Gas Reporting Program (GHGRP) which is a part of their Office of Atmospheric Protection (OAP). The different types of emissions being tracked are dictated by the subparts of the ruling (codified at 40 CFR Part 98 for those interested), in the case of CCUS, the three most relevant subparts are subpart PP, UU, and RR.



The image to the left, courtesy of the GHGRP’s “Supply, Underground Injection, and Geologic Sequestration of Carbon Dioxide” page, Subpart PP involves the supply of carbon dioxide, not the storage of it, so it was not included in this project.

Subpart UU covers underground injection whereas subpart RR covers geologic sequestration. Underground injection is self-explanatory, and there are a variety of reasons why a company may inject carbon. One of the most common reasons is enhanced oil recovery (EOR), which “is a process for extracting oil that has not already been retrieved through the primary or secondary oil recovery techniques” (“Enhanced Oil Recovery”). EOR is commonly used by oil companies, such as ExxonMobil, and is a fairly common use of carbon injection (per the red squares on the map). On the other hand, geologic sequestration (GS), which is covered by subpart RR, is “the process of storing carbon dioxide (CO2) in underground geologic formations” ("What's the Difference"). This process can also be used for EOR, so there is some overlap between subparts UU and RR.

Background

Inflation Reduction Act and the Biden Administration

The Inflation Reduction Act of 2022 (IRA) was signed into effect by President Biden on August 16, 2022. The act was aimed at increasing tax breaks for the purpose of "boost[ing] clean energy, reduc[ing] healthcare costs, and increas[ing] tax revenues" (McKinsey & Company). Specifically, the IRA expands on the existing 45Q tax credit which is a "performance-based tax credit for carbon management projects" and it raises the credit for sequestered carbon dioxide from $60/tonnes to $85/tonnes (Carbon Capture Coalition). The increase in this tax credit is to encourage more people to invest in or participate in carbon storage projects.

Health and Safety Concerns

Where can you build a carbon sequestration facility? In pretty much any middle to high income neighborhood, the thought of a carbon sequestration facility, or anything related is impossible. A company wouldn't even dream of pitching that internally, nevermind actually to the community itself. Furthermore, middle and high income neighborhoods often have strict zoning laws that protect the sanctity of the community. With roots in segregation, this is not true for lower income neighborhoods. That's the reason why neighborhoods like St Rose, Louisiana obtain nicknames like "Cancer Alley".

St Charles Clean Fuels (SCCF) is planning to build a blue ammonia and hydrogen facility nearby St Rose where they will capture and release carbon dioxide to a later sequester. Critics, of course, not only doubt the industry's 90% carbon dioxide capture claims, but point out that these facilities release more than just carbon dioxide, which is left to pollute the local air.

Furthermore, to transport the carbon dioxide from this facility to sequestration facilities, they will have to travel thousands of miles via pipelines built throughout the United States.

Though the events that occurred in Satartia, MS happened more than four years ago, it returned to the public consciousness in the wake of the Biden administration's investment in carbon dioxide storage and transportation projects.

In February 2020, A pipeline transporting carbon dioxide between facilities experienced a rupture which released carbon dioxide into the air for around four hours. Even though the rupture occurred in open air, the carbon dioxide did not disperse enough to avoid serious injury to the residents. More than 200 people from the town were evacuated, while 45 were admitted to the hospital. Many of Satartia's residents still suffer from impairments that they attribute to this incident. Federal regulators that investigated the pipeline rupture found that the operator, Denbury

Inc., violated federal regulations which led to the leak. Despite knowing immediately that the leak had occurred, the company did not alert the local authorities.

This incident occurred at a pipeline, and not at a carbon injection site, but this story is still a relevant concern. More carbon dioxide capture and storage projects means more pipelines to transfer carbon dioxide between sites. Jesse Jenkins, a professor at Princeton University, predicts that the amount of pipelines in America could grow from 5,300 miles (at the time of the article's writing) to more than 65,000 miles. Competitors of Denbury claim that the fault lies with the company, and that regulation surrounding pipelines is strong, but without strong enforcement and preventative measures, similar incidents as what happened in Satartia, MS are bound to happen and affect others.

Denbury Inc's punishment was a $2.8 million fine to the US Government.

Methodology

Research Question

For this capstone project, my research question was: Who makes up the communities surrounding carbon capture facilities reporting to the EPA's GHGRP? I believe that descriptive analysis achieved my goal better, since there is no merit to predicting who lives near facilities, since they are located in a variety of types of locations (rural and urban). For the purpose of data analysis, I defined the "community" of a facility as the area within a 20-mile radius around the facility.

My approach had two phases: a data collection and curation phase, and a user interface creation phase. All of my data collection, modification, and analysis happened using Quarto and Shiny in R.

FLIGHT Data

The first dataframe that I downloaded was from the EPA's GHGRP's FLIGHT (Facility Level Information on GreenHouse gases Tool) from which you can find the reported emissions and facility data. I specifically downloaded the subpart RR data and the subpart UU data.

	Facility Id	FRS Id	Facility Name	City	State	Zip Code	Address	County	Latitude	Longitude	Primary NAICS Code	Industry Type (subparts)	Total Mass of CO2 Sequestered
1	1013701	110070931003	30-30 Gas Plant	Plains	TX	79355	2300 FM 1622	YOAKUM COUNTY	33.05188	-102.88792	211130	C,PP,RR (RPT),W-PROC	12354.6
2	1005661	110050297936	Archer Daniels Midland Co.	DECATUR	IL	62521	4666 FARIES PARKWAY	Macon	39.86750	-88.88500	311221	C,II,PP,RR (RPT)	428580.4
3	1013609	110070385361	Campo Viejo Gas Processing Plant	Plains	TX	79355	1548 County Road 165	YOAKUM COUNTY	33.14992	-102.99117	211130	C,PP,RR (RPT),W-PROC	76658.3
4	1010117	NA	Core Energy Otsego County EOR Operations	Gaylord	MI	49735	597 Kubacki Road	NA	45.03384	-84.51147	211120	RR (RPT),W-ONSH	311307.6
5	1011767	110070082145	Denver Unit	Denver City	TX	79323	2611 State Hwy 214	YOAKUM COUNTY	33.00338	-102.81901	211120	RR (RPT)	2849399.5
6	1009999	110071159834	Farnsworth Unit CO2 Flood	Farnsworth	TX	79033	NA	OCHILTREE COUNTY	36.26530	-101.02600	211120	RR (RPT)	92201.1
7	1002440	110070292846	Great Plains Synfuels Plant	BEULAH	ND	58523	420 COUNTY 26	MERCER COUNTY	47.36081	-101.83822	221210	C,G,LL,PP,RR (RPT)	NA
8	1012121	110071162318	Hobbs Field	Hobbs	NM	88240	1017 West Stanolind Road	LEA COUNTY	32.68204	-103.14912	211120	RR (RPT)	2276827.9
9	1010975	110071159322	North Burbank Unit	Webb City	OK	74652	NA	OSAGE COUNTY	36.82491	-96.73257	211120	RR (RPT)	652430.2
10	1013810	NA	Petra Nova West Ranch	Vanderbilt	TX	77991	1421 Mobil Oil Road	NA	28.80207	-96.62646	211120	RR (RPT)	-16814.5
11	1001157	110027922602	RED TRAIL ENERGY, LLC	RICHARDTON	ND	58652	3682 HWY 8 SOUTH	STARK COUNTY	46.87864	-102.29604	325193	C,PP,RR (RPT)	81963.8
12	1011064	110057067739	Red Hills Gas Processing Plant	Jal	NM	88252	1934 W NM Highway 128	LEA COUNTY	32.21043	-103.52141	211130	C,PP,RR (RPT),W-PROC	23775.0
13	1002150	110071162292	Shute Creek Facility	KEMMERER	WY	83101	NA	LINCOLN COUNTY	41.88700	-110.09446	211130	C,PP,RR (RPT),W-PROC	395332.2
14	1013793	110071232134	West Seminole San Andres Unit	Seminole	TX	79360	100 NW 7th Street	GAINES COUNTY	32.72136	-102.65248	211120	RR (RPT)	768740.6

A screenshot of the subpart RR data taken in RStudio

	Facility Id	FRS Id	Facility Name	City	State	Zip Code	Address	County	Latitude	Longitude	Primary NAICS Code	Industry Type (subparts)	Quantity of CO2 Received for Injection
1	1002515	110043804096	ARTESIA GAS PLANT	ARTESIA	NM	88210	NA	EDDY COUNTY	32.75640	-104.21110	211130	C,PP,UU,W-PROC	confidential
2	1010508	110071159820	Adair San Andres CO2 Injection Unit - Permian Basin 430	Houston	TX	77056	2000 Post Oak Blvd, Suite 100	HOCKLEY COUNTY	33.47193	-102.53186	211120	UU	confidential
3	1009998	110071160568	Albert Spicer Upper Morrow Unit	Booker	TX	79005	NA	OCHILTREE COUNTY	36.47250	-100.55170	211120	UU	confidential
4	1007367	110071160535	BPE GPRP Grasslands Gas Plant	Cartwright	ND	58838	NA	RICHLAND COUNTY	47.59043	-104.00050	211130	C,NN-FRAC,PP,UU,W-PROC	confidential
5	1011121	110058238991	Bell Creek EOR Facility	Belle Creek	MT	59317	2 Encore Road	POWDER RIVER COUNTY	45.35459	-105.67196	211120	UU	confidential
6	1011485	110071160693	Big Sand Draw CO2 Facility	Riverton	WY	82501	NA	OKLAHOMA COUNTY	42.75407	-108.17038	211120	UU	confidential
7	1009946	110071159434	Booker Field Area	Booker	TX	79005	NA	OCHILTREE COUNTY	36.45800	-100.54620	211120	UU	confidential
8	1009630	110071160367	Brookhaven EOR Facility	Brookhaven	MS	39601	1030 California Road	LINCOLN COUNTY	31.59027	-90.50485	211120	UU	confidential
9	1014609	110071397381	CHSU Miller EOR Facility	Marmarth	ND	58643	NA	BOWMAN COUNTY	46.21250	-103.94120	211120	UU	confidential
10	1009997	110071160567	Camrick Unit	Balko	OK	73931	NA	BEAVER COUNTY	36.52100	-100.89750	211120	UU	confidential
11	1009631	110055517310	Cranfield EOR Facility	Cranfield	MS	39661	887 B Hwy 84	FRANKLIN COUNTY	31.48664	-91.09694	211120	UU	confidential
12	1012519	110070082340	Delaware Basin Gas Plant	Coyanosa	TX	79730	1220 Conty Road 101	NA	31.24076	-103.06533	211130	C,PP,UU,W-PROC	confidential

A screenshot of the first 12 rows of the subpart UU data taken in RStudio

The above two screenshots show what the FLIGHT data looks like for subpart RR and UU. They have the same columns except for the last column which reports the amount of CO2 that was sequestered for subpart RR, and the amount of CO2 that was received for subpart UU. The most important columns shown here are the 'Facility Id', which was as a unique identifier for each facility, and the 'Latitude' and 'Longitude' columns. The two datasets were subsetted to just those three columns.

Google Places API and Dataframe

From the coordinates of the facilities, I used the Google Places API in R to generate a

Medical Services	Hospital, Pharmacy, Health, Doctor
Schools	School, Secondary School, University
Transportation	Transit Station, Train Station, Light Rail Station, Gas Station, Airport
Social Locations	Cafe, Restaurant, Gym, Club, Bar, Zoo, Laundry, Library, RV park, Park, Lodging, Campground
Essential Services	Bank, City Hall, Courthouse, Police, Post Office, Fire Station
Places of Worship	Synagogue, Church, Mosque, Place of Worship, Hindu Temple
Stores	Bakery, Shopping Mall, Supermarket, Store, Convenience Store, Department Store

dataframe of different types of locations of interest that surround the facilities. The table below presents two columns: the left column lists the broad categories, and the right column specifies the types of locations in Google Places that were collected under each broad category.

The two broad categories I want to further explore are the "Social Locations" and "Essential Services". My operational definition of "Essential Services" were locations that are vital to keeping order in a city or town. In particular, I was interested in whether there were police and fire stations within the radius of facilities. When it comes to accidents happening at these facilities, being within range of such emergency services would make mitigating the situation easier. As for social locations, my operational definition was locations where people would *choose* to gather. That is why a campground is included, even though it may not necessarily be seen as a social location, it is a place where people would choose to go to, and where large, diverse groups of people could gather.

```
{r}
for (a in 1:nrow(seq)) {
  site <- seq[a,]

  category <- c("hospital","pharmacy","health","doctor")
  medical <- data.frame(
    lat = numeric(),
    lng = numeric(),
    type = character(),
    distance = numeric()
  )

  for (i in category) {
    request <- google_places(
      location=c(site$Latitude,site$Longitude),
      rankby='distance',
      place_type=i)
    medical <- rbind(request$results$geometry$location,medical)

    while (exists(addQuotes(request$next_place_token))) {
      request <- google_places(
      location=c(site$Latitude,site$Longitude),
      rankby='distance',
      place_type=i,
      page_token=request$next_page_token)

      medical <- rbind(request$results$geometry$location,medical)
    }
  }

  if (nrow(medical)>0) {
    medical$type <- c("Medical")
    medical$distance <- distHaversine(cbind(medical$lng,medical$lat),
                                      cbind(site$Longitude,site$Latitude))
    colnames(medical) <- c("lat","lng","type","distance")
    medical <- medical[!duplicated(medical),]
    medical$site <- site$`Facility Id`
    medical <- subset(medical,medical$distance<=r)
    locations <- rbind(medical,locations)
  }
}
```

The above chunk shows the basic format that I used to generate the Google Places dataset. The code runs for each individual facility. First, I establish the subcategories that I was the API to look for, in this case it is looking for any location tagged as "hospital", "pharmacy", "health", and "doctor". Then, I create an empty dataframe to collect the coordinates, location type, and distance from the facility coordinate for the facility. Then, for every facility, the API looks for all the locations of the specified type surrounding the locations. Google Places' API can only show 20 results at a time, so the code repeats the call three times to generate the maximum 60 results. The API works by looking at the closest locations and progressing outwards until it reaches the maximum number of results.

After the API generates the data, I label the data type, "Medical" in this case, and use the *distHaversine* formula from the *geosphere* package, which calculates the distance between the facility location coordinates and the generated location coordinates. I then subset the resulting data by 'r', which is 20 miles. The empty dataframe that was created early is then copied to the master locations dataframe and the data frame is rewritten so that the process can happen for another facility.

The API runs for every facility for every category, resulting in the following data frame:

	lat	lng	type	distance	site	state	latitude	longitude
1	46.88474	-102.3173	Essential Services	1753.6838	1001157	North Dakota	46.87864	-102.2960
2	46.90088	-102.4231	Essential Services	9979.7847	1001157	North Dakota	46.87864	-102.2960
3	46.90218	-102.0466	Essential Services	19154.2698	1001157	North Dakota	46.87864	-102.2960
4	46.88424	-102.3144	Essential Services	1527.8609	1001157	North Dakota	46.87864	-102.2960
5	46.90088	-102.4242	Essential Services	10057.8931	1001157	North Dakota	46.87864	-102.2960

The total final locations data frame has 27,145 rows.

Zip Code and the ACS Data

The second dataset that was used for analysis was the American Census Survey which is conducted by the United States Census Bureau. I downloaded this data directly from the Census Bureau's website and modified it in Quarto to better fit my needs. In a previous run of a similar project, I had used the county that a facility was in to represent the population in the community, but I was concerned that if a facility was located on the very edge of a county that this would not be a good representation. Thus, I chose to calculate all of the zipcodes that are a part of the 20 mile surrounding area of the facility and used their average as a representation of the community's demographics.

R has a library, *zipcodeR*, that has a function that can find all of the zip codes within a specified radius of a set of coordinates. The chunk below displays the code that I used to generate the dataframe of all of the zipcodes pertaining to a facility.

```
{r}
zip <- data.frame(
  zipcode = numeric(),
  distance = numeric(),
  id = factor()
)

library("zipcodeR")

for (i in 1:nrow(seq)) {
  zip_hold <- search_radius(seq$Latitude[i],seq$Longitude[i],20)
  zip_hold$id <- seq$`Facility Id`[i]

  zip <- rbind(zip_hold,zip)
}
```

The image below shows a snapshot of the ACS Census Bureau data for age demographics in the United States, by zipcode. This dataset had 33,775 rows and 459 columns.

	GEO_ID	NAME	S0101_C01_001E	S0101_C01_001M	S0101_C01_002E
1	Geography	Geographic Area Name	Estimate!!Total!!Total population	Margin of Error!!Total!!Total population	Estimate!!Total!!Total population!!AGE!!Under 5 years
2	860Z200US00601	ZCTA5 00601	16834	506	593
3	860Z200US00602	ZCTA5 00602	37642	205	1194
4	860Z200US00603	ZCTA5 00603	49075	963	1659
5	860Z200US00606	ZCTA5 00606	5590	264	204
6	860Z200US00610	ZCTA5 00610	25542	344	762
7	860Z200US00611	ZCTA5 00611	1315	382	77
8	860Z200US00612	ZCTA5 00612	63312	1805	2418
9	860Z200US00616	ZCTA5 00616	9625	1319	242
10	860Z200US00617	ZCTA5 00617	22573	241	797
11	860Z200US00622	ZCTA5 00622	7577	979	235
12	860Z200US00623	ZCTA5 00623	39406	979	1102
13	860Z200US00624	ZCTA5 00624	21648	516	1138

The datasets included a lot of duplicate information, such as many different ranges of ages. In the snapshot, you can see that the fourth column, S0101_C01_001M is the margin of error. Every category is duplicated to include the margin of error. The age ranges were small, so I combined them to create four age categories: under 5 years old, between ages 5 and 19, ages 19 to 64, and ages 64 and up. I initially wanted the age range to be 5 to 18 or 5 to 17. However, since the available data only included the age range of 15 to 19, it was not possible to adjust it.

A similar process occurred for the ACS income data. Many of the ranges were overly small, so I combined them to create seven income ranges: less than $25,000 a year, between $25,000 and $50,000 a year, between $50,000 and $75,000 a year, between $75,000 and $99,000 a year, between $100,000 and $150,000 a year, between $150,000 and $200,000, and more than $200,000 a year.

For the ACS ethnicity data, I used the given categories which were Asian, Black (or African American), Native (American Indian or Alaska Native), Pacific Islander (Native Hawaiian or Other Pacific Islander), White, multiple ethnicities, and Other.

Using the total population, I converted all of the categories into percentages, joined the data with the zip code data frame and then averaged the percentages by facility. The image below shows the resulting data frame for the ethnicity data.

	id	totalpop	white	black	native	asian	pacific	other	multiple
1	1001157	678.5000	0.9239830	0.0072463768	2.056192e-02	0.0070656710	3.875969e-03	0.0046520416	0.032615024
2	1002150	1321.6667	0.9414849	0.0006287331	3.221626e-03	0.0051879213	0.000000e+00	0.0113463915	0.038130443
3	1002440	1575.6000	0.9180718	0.0012319064	5.383819e-02	0.0080970015	0.000000e+00	0.0080073914	0.010753688
4	1002515	6217.6667	0.9043410	0.0011319534	7.761966e-03	0.0019404916	0.000000e+00	0.0314431508	0.053381486
5	1003698	187.0000	1.0000000	0.0000000000	0.000000e+00	0.0000000000	0.000000e+00	0.0000000000	0.000000000
6	1003719	551.8571	0.9242694	0.0135664363	8.008500e-03	0.0229575029	0.000000e+00	0.0101532992	0.021044877
7	1003767	2218.0000	0.4997817	0.0191153788	8.358028e-04	0.0119952891	4.558924e-04	0.4190138071	0.048802143
8	1004617	13890.5000	0.9171688	0.0095262151	4.708981e-03	0.0054126222	3.608415e-05	0.0257279977	0.037419262
9	1004893	13164.5000	0.6487277	0.0145671276	1.192092e-02	0.0023648324	0.000000e+00	0.0401195072	0.282299904
10	1004924	2393.3333	0.7741994	0.0086258897	2.888902e-03	0.0065098688	0.000000e+00	0.0240514466	0.183724444
11	1005641	16147.7500	0.6365297	0.0225842911	6.914836e-03	0.0031068637	0.000000e+00	0.0602027437	0.270661566

Limitations

One glaring limitation to this project is the choice of a 20 mile radius. 20 miles is a huge distance. For perspective, the Dulles Airport (Dulles, VA) is 21 miles from Foggy Bottom, DC. Generalizing an entire community by a radius that large loses the nuances of the community. A ten mile radius may have slightly different age demographics than the 20 mile radius, for example, and when it comes to an area with a very small population, that is a significant change.

Google Places API

My definition of the categories was very subjective. I think I could have also been more specific to create a more holistic understanding of the community. Especially when it comes to the "Essential Services" category, I believe that I could have separated it into "Emergency Services" (for the fire station and police categories) and into another category for government related services (for banks and courthouse, for example).

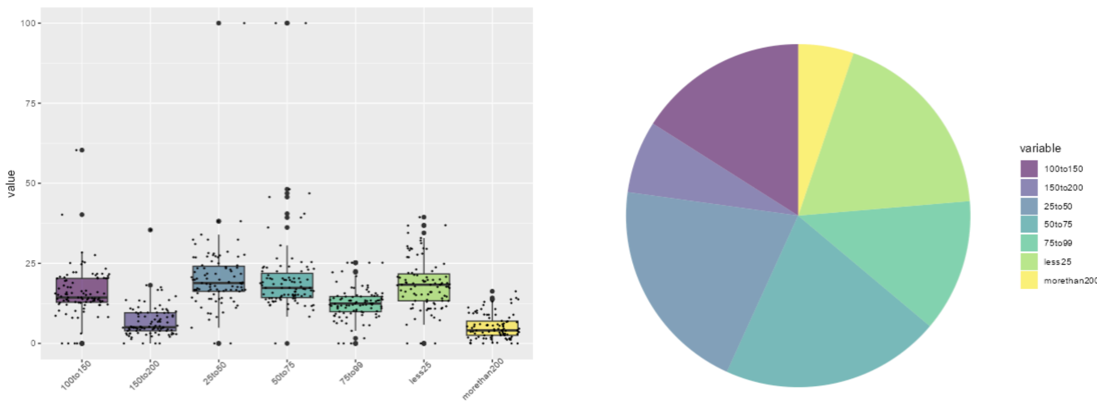
The Google Places API maxes out at 60 search results, so for each category, you can only generate up to 60 locations. While this may be fine for less populous areas, it may give a fractured image of the communities for more urban areas. Since I was not really able to look into each of the 98 facilities individually, I was not able to confirm that the Google Places API did not max out and leave out vital locations.

American Census Survey

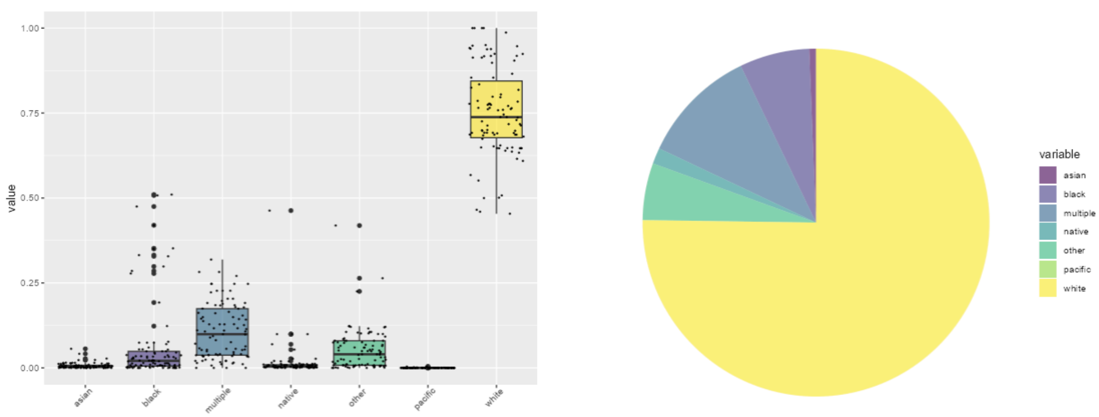
The survey data is from 2022, so it may not fully reflect the current demographics of the community. Additionally, while using zip codes effectively covers the 20-mile radius, it also extends beyond that boundary. This creates uncertainty about whether some included zip codes encompass areas with significantly different populations, which could impact the accuracy of the resulting data.

Results

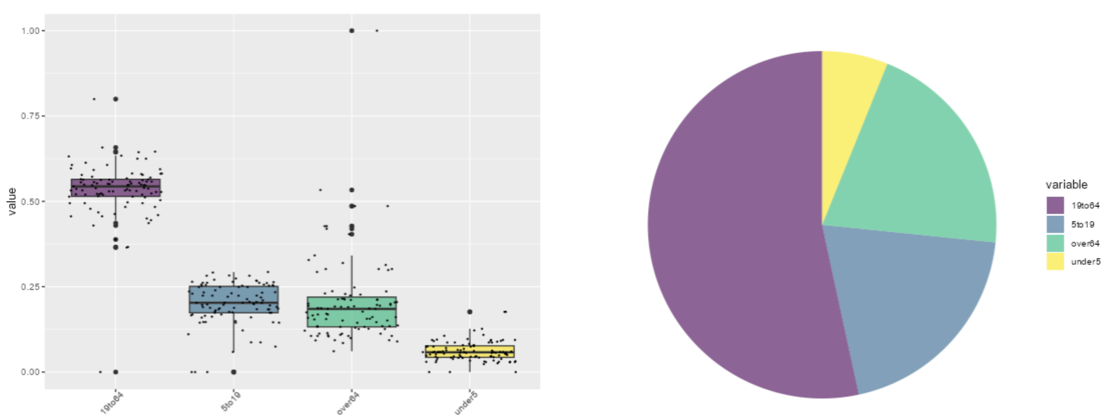
Data Analysis



The income demographics of the facilities

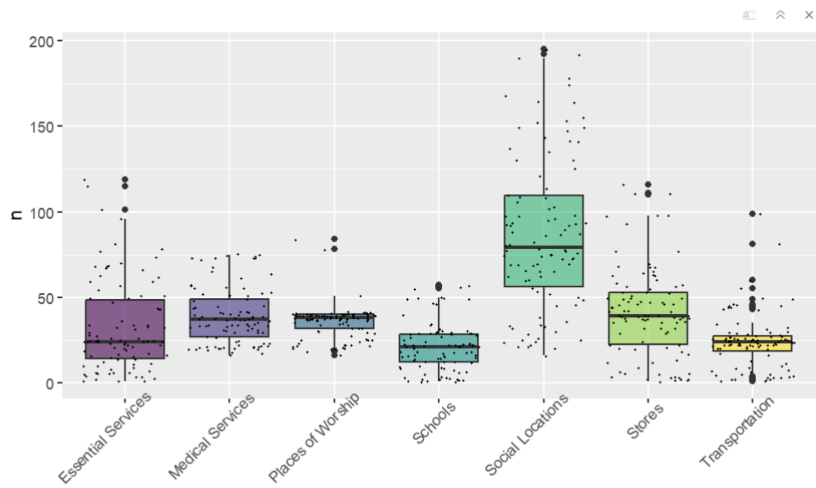


The ethnicity demographics of the facilities



The age demographics of the facilities

The graph above shows the distribution of age for all facility communities. As shown by the pie chart, and predictably, the population is largely made up of individuals between the ages of 19 and 64. I don't think that any of the census demographic charts indicate that there is any significant deviation from the general US population.



The results of the Google Places API are the most interesting. The thick middle line within each box shows the median of the points, while the top and bottom line shows the range. The three location categories of which there was not a single facility that had none of were Medical Services, Places of Worship, and Social Locations. These are all within expectations, especially if you consider that many rural areas are made of Christains. Social locations are located everywhere, and medical services *should* be within a 20 mile radius of any location.

I was stunned that there were facilities that had no essential services, meaning there were no police stations, no fire stations, no banks, etc. within a 20 mile radius of these locations. While this also has some dangerous implications should there be any complications at the facility, I also think it's curious that every facility has at least 20 social locations, but none of these essential locations.

I think it is also interesting to see just how populous these communities are. As I've pointed out previously, there are lots of social locations surrounding these facilities. The box plot is scaled specifically for the Social Locations category since it is the most far reaching. There can be upwards towards 200 social locations, if not more, since the API maxes out at 60 location results per category.

The ranges of these box plots also indicates that there are huge differences in population densities between facilities. Some facilities are located in significantly less populated areas than others. This is also something we know anecdotally, since there are some facilities located in or near mid-sized cities and other that are located in extremely rural areas.

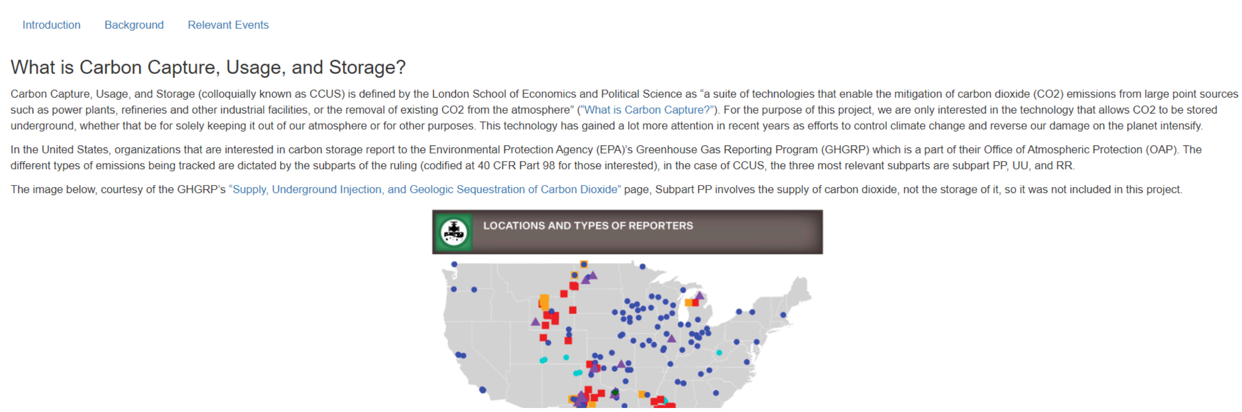
In conclusion, the results from the Google Places API highlight significant contrasts in the types and distribution of locations surrounding carbon injection facilities. The consistent presence of Medical Services, Places of Worship, and Social Locations aligns with expectations, reflecting essential and social infrastructure in rural and urban communities. However, the absence of essential services such as police stations, fire stations, or banks within a 20 mile radius of some facilities raises critical safety concerns, especially given the potential risk of complications at these facilities.

These findings emphasize the importance of tailored safety regulations and emergency preparedness to address the unique needs of each community, ensuring that all residents, regardless of location, are adequately protected.

User Interface

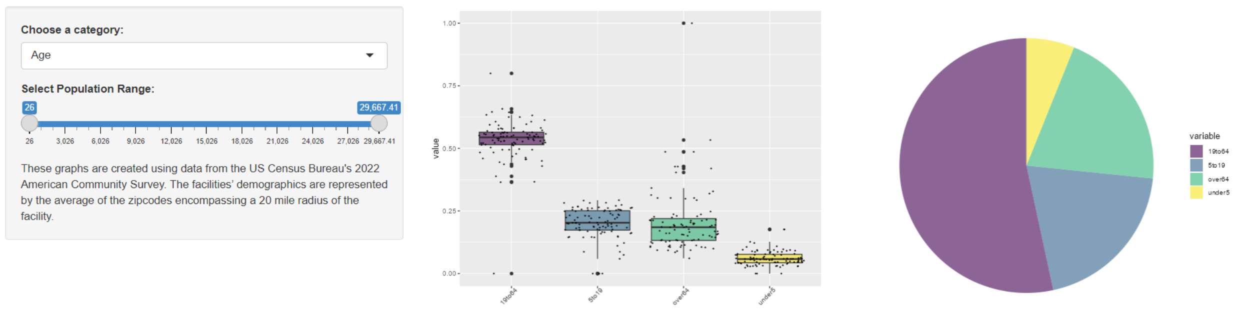
The user interface that I created has four pages: Overview, Census Demographics, Facility Locations, and Individual Sites.

The Overview page includes all of the information in the report's introduction section. This page splits the introductory information into three sections: an Introduction, the Background, and Relevant Events. The introduction part focuses on introducing the base topic of carbon capture and the Greenhouse Gas Reporting Branch. The background section focuses more on introducing my motivations for this project, as well as some relevant history for why understanding the communities surrounding these facilities is important. Finally, the relevant events section offers further reading into events that are related to my project and this topic.



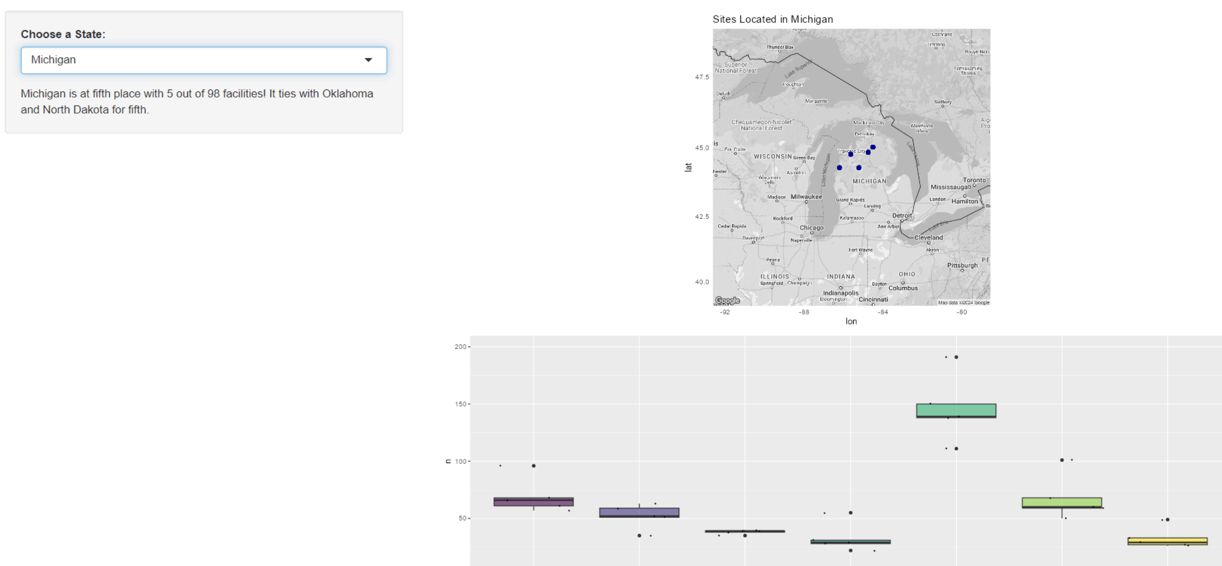
The Introduction section of the Overview page

The Census Demographics page allows the user to explore the three types of demographic data that was collected for this project: Age, Income, and Ethnicity. Changing the category updates the box plot and pie chart, both generated using the *ggplot* package. In the side panel, there is also the option to change the population range. Though my project did not explore it deeply, I still believe that population must have an impact on the demographics of a community. This option allows the user to explore this idea as well.



The Census Demographics page, selecting the 'Age' category, and looking at the demographics for all populations

Third is the Facility Locations page. This page gives a more holistic view of where facilities are located as it relates to the state it is located in. This page includes a side panel where the user can change the selected state, a map that plots all of the facilities located in that state, and a box plot that shows the distribution of the Google Places locations for the state.



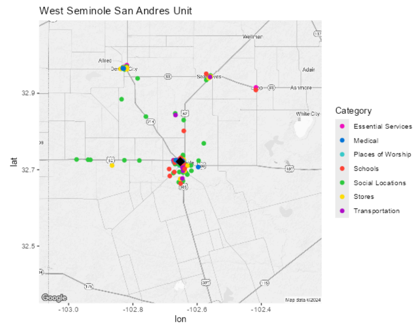
The Facility Location page set to Michigan, not pictured is a table explaining the location types in the box plot, which is below the plot

Lastly is the Individual Sites page. This page allows the user to view a map displaying the locations to each of the Google Places location types as it relates to the facility, represented by the big black diamond at the center of the map. I also included a table so that the user can see specifically how many locations there were per category.

Choose a Site:

1013793

Facility Id	1013793
Facility Name	West Seminole San Andres Unit
City	Seminole
State	TX
Industry Type (subparts)	RR (RPT)



type	n
Essential Services	31
Medical	31
Places of Worship	38
Schools	21
Social Locations	99
Stores	52
Transportation	27

Location Type	Description
Essential Services	Locations providing key public services, such as banks, city halls, courthouses, police stations, post offices, and fire stations. These services are critical for community safety, legal functions, and financial transactions.
Medical	Locations offering healthcare-related services, such as hospitals, pharmacies, general health centers, and doctor's offices, focused on patient care and medical treatment.
Places of Worship	Religious sites for prayer, worship, and spiritual activities, including synagogues, churches, mosques, Hindu temples, and other places of worship for various faiths.
Schools	Educational institutions providing a range of academic services, including primary, secondary, and higher education, such as schools, secondary schools, and universities.
Social Locations	Places designed for socializing, relaxation, and recreation, such as cafes, restaurants, gyms, clubs, bars, zoos, libraries, and parks. This category includes both public and private spaces for leisure activities.
Stores	Commercial establishments where goods are sold to the public, including bakeries, shopping malls, supermarkets, convenience stores,

Conclusion

Based on the Census and Google Maps data collected, carbon injection facilities do not appear to be disproportionately located in marginalized areas. However, considering past accidents, it is essential for regulators to enforce strict compliance with safety regulations to protect all communities. CCUS projects pose unique threats to both low and high population areas. In sparsely populated areas, emergency services may lack the capacity to respond effectively to leaks or accidents, putting residents at greater risk. In densely populated areas, the potential impact is magnified due to the larger number of residents who could be affected.

Even if these areas are not marginalized in the evaluated ways, this project highlights the vibrancy and diversity of the communities surrounding these facilities. These populous and colorful neighborhoods deserve robust protections. Moving forward, prioritizing strong regulation, comprehensive safety measures, and emergency preparedness is critical to safeguarding public health and ensuring the well-being of all residents.

References

Carbon Capture Coalition. 45Q Primer. Carbon Capture Coalition, Nov. 2023. https://carboncapturecoalition.org/wp-content/uploads/2023/11/45Q-primer-Carbon-Capture-Coalition.pdf.

"Enhanced Oil Recovery." Investopedia, Dotdash Meredith. https://www.investopedia.com/terms/e/enhanced-oil-recovery.asp. Accessed 8 Dec. 2024.

McKinsey & Company. "The Inflation Reduction Act: Here's What's in It." McKinsey & Company, 17 Aug. 2022. https://www.mckinsey.com/industries/public-sector/our-insights/the-inflation-reduction-act-heres-whats-in-it.

"What's the Difference Between Geologic and Biologic Carbon Sequestration?" U.S. Geological Survey. https://www.usgs.gov/faqs/whats-difference-between-geologic-and-biologic-carbon-sequestration#:~:text=Geologic%20carbon%20sequestration%20is%20the,rock%20formations%20in%20geologic%20basins. Accessed 8 Dec. 2024.

"What Is Carbon Capture and Storage, and What Role Can It Play in Tackling Climate Change?" Grantham Research Institute on Climate Change and the Environment, London School of Economics and Political Science. https://www.lse.ac.uk/granthaminstitute/explainers/what-is-carbon-capture-and-storage-and-what-role-can-it-play-in-tackling-climate-change/#:~:text=What%20is%20carbon%20capture%2C%20usage%20and%20storage%20(CCUS)%3F,CO2%20from%20the%20atmosphere. Accessed 8 Dec. 2024.