Legacy Pollution – Hazardous Waste Sites README

Category	Information
High-level overview	Georgia sites classified as hazardous waste facilities (treatment, storage, and disposal facilities and large quantity generators) by the EPA as of 2020. Data set contains 8 columns to describe name and location of these sites, including Handler ID, County, and Latitude/Longitude.
Search keywords	Legacy pollution, hazardous sites inventory, EPA RCRAInfo Search, Climate and Economic Screening Tool (CEJST)
Data Dictionary? Y/N	N
Does it contain PII? If yes – STOP HERE	No PII
Use cases. **Links to example uses of a few sentences for intended use**	This data was used to create the Legacy Pollution category of the CEJST, an interactive geospatial map that highlights areas across the U.S. that are disadvantaged: https://screeningtool.geoplatform.gov/en/#3.02/33.55/-97.56 In Fall 2023, Science for GA and an Emory Data Think group used the data to examine the impact of water pollution on local Georgia communities.
How to access the Raw Data Source **(Link to data source and access instructions)**	EPA Treatment, Storage, and Disposal Facilities (TSDF) from 2020, calculated from EPA's Resource Conservation & Recovery Act (RCRA) database as compiled by EPA's EJScreen • RCRAInfo Search can be filtered by state • Public and consistent nationwide Data as described in CEJST methodology: https://screeningtool.geoplatform.gov/en/methodology#prox-haz
Dynamic or static **(Does the analyst have to re-download and re-clean, or is the data set up-to-date)**	Dynamic – data set is up-to-date for defined place and time period (GA, 2020). Analyst may need to re-clean data for their specific project purposes
How is the raw data maintained?	Raw data is maintained in by the Environmental Protection Agency (EPA)
Geographic and time coverage of data.	Data set covers census tracts in Georgia as of 2020 Raw data available for all 50 states, D.C., and U.S. Territories in RCRA database

Caveats or limitations on data.	Data set may not be updated with new information post-2020 and uses census tract boundaries from 2010.
When was the raw data accessed for "cleaning"?	Fall 2023
How was the data verified?	Raw data selected for use by the Council on Environmental Quality (CEQ). Collected and maintained by U.S. EPA Dataset compared to Georgia Environmental Protection Division (EPD) July 2023 Hazardous Site Inventory
Link to code and instructions for "cleaning"	https://github.com/sci4ga/ej-bucket-exploration/tree/main/DataThink/Legacy%20Pollution (see "Sites.R")
Completeness vs. sparsity	More complete than sparse
Link to any code and instructions for data analysis.	https://github.com/sci4ga/ej-bucket-exploration/tree/main/DataThink/Legacy%20Pollution (See "Legacy Pollution.ipynb")
Link to cleaned dataset	Data Think Fall 23 – Legacy Pollution Datasets
Link to GitHub Repository for READMEs, code, etc	GitHub Repo – Data Think 23 Legacy Pollution (See "Hazardous Waste Sites" CSV and Code Book)