

Table of Contents [Google Drive Link](#)

Table of Contents	1
Introduction to Database	1
Dataset File Types	1
File Conversion Tools	2
Navigating the Dataset	2
Downloading from Dataset -	2
Possible Scientific Questions and Visualizations	
Examples of Questions and Visualizations	2
FAQ section	3
Version history	3

Introduction

This database is accessed using this following link: [Dataset - Catalog](#)

This dataset consists of data that pertains to climate. In this dataset you may find anything from aquaculture data to United States annual average precipitation. It is created by filtering all the data.gov datasets by climate. Using this dataset, you can create visualizations about the climate in America. The data is collected by the United States federal government as well as its particular departments such as the EPA and the Food and Nutrition Service. The data is collected in the United States and is taken from over 40 states. You will find data ranging from federal datasets to statewide datasets.

The data collection was created using the topic “climate”. The region was also made “United States” in order to limit the data. Without this limit, the collection had an overwhelming amount of datasets. The location can easily be changed with the dropdown menu.

Topics:

Climate x

Filter by location

Clear

United States

▼

145 datasets found

By creating visualizations of the different datasets, you may find trends over time that could predict key indicators of climate change. You may see certain trends that may be useful for research questions. Some examples of these questions and visualizations are included later in the document and can also be reached in the table of contents. The data collection has different aspects of the environment from soil data to air and water data. Anything to do with climate pertaining to the United States can be mostly found in this collection. To reiterate, you can use the “Search datasets” bar to find any other criteria you may be looking for. You can also search variations of “climate” such as “climate change”, “environment”, etc. These new data collections from the new filters may be broader and thus larger and harder to navigate. You can also use specific tags such as “air data”, “soil data”, etc.

Some limitations to this dataset include it being only 145 datasets. By filtering using the keyword, “climate”, you might be missing some datasets that pertain to climate but were not labeled properly in the system. We also add that this is only data collected by the United States government and its entities and so the data may be limited in that aspect.

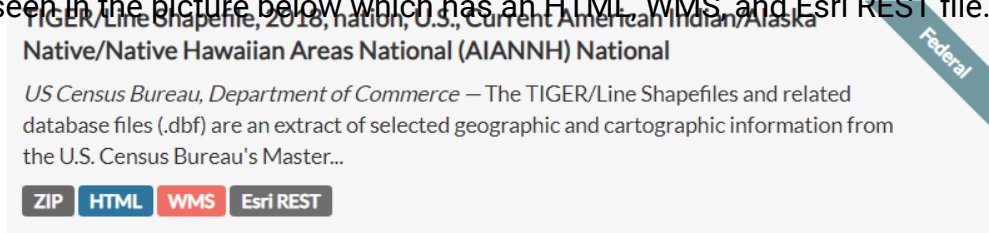
Navigating through the rest of this document, you will learn more about the database and the file types you may use when downloading and interacting with the data. You will also find an FAQ section that may answer any questions you may have about this data collection.

Dataset File Types

The main file types found in our data set (in order of most common to least common):

- Esri REST
 - REST stands for REpresentational State Transfer which is a file type based on HTTP protocols.
 - It is primarily used to project and manipulate maps or any other kind of geographically based data.
- HTML
 - Used to create and structure documents found in a web browser.
 - In this dataset, HTML files will redirect you to a website which will carry the climate data you are looking for.
- XLS/XLSX
 - This file type can be manipulated in Microsoft Excel to perform both basic and complex statistical and data analysis.
 - Can contain thousands of cells of data on various climate statistics.
- WMS
 - Stands for Web Map Service.
 - Used to portray georeferenced map data over the internet which commonly uses GIS(Geographic information system) files.
- CSV
 - Comma Separated Value files are very versatile and can be used on many different software to manipulate or display data.
 - Can be imported into Excel or used in coding software such as C, Python, or SQL to name a few.

- Many datasets found contain multiple file types for a single topic. This can be seen in the picture below which has an HTML, WMS, and Esri REST file.



File Conversion Tools

- File conversion refers to the transfer of a file from one format to another which can be done in order for your data to be manipulated by different programs.
- One easy method of converting your data to and from an Excel file is through Excel's built in file conversion system.
 - To import HTML, XML, or text scripts into an excel file: Go to the **Data tab**, select **Get & Transform Data group**, and from there click the file type you are using.
 - Additionally you can use this method to convert your Excel file to a HTML, XML, or text script.
- Another way of converting files is through file conversion websites.
 - [CloudConvert](#) is a very useful conversion tool that can convert over 200 different file formats for free. (It will convert all file types found in this dataset)

Navigating the Dataset

- Different types of filtering methods can be used
 - Location
 - Able to filter location by zooming in and out of the given map
 - Able to search a location by typing at least 3 related letters



- Climate

- Users are able to choose from a variety of different options
- Numbers next to each category will narrow down the datasets into that number
- Users are able to choose multiple categories to narrow down their search

The screenshot displays the EnviroAtlas search results interface. On the left, a sidebar contains a 'Filter by location' section with 'United States' selected, and a 'Topics' section with 'Climate' (38) and 'Water' (38) selected. Below these are 'Topic Categories' including 'Atmospheric, Earth...' (3), 'Ecosystem Services...' (3), 'Ecosystem Vulnerability' (2), 'Human Health' (2), 'Hydrology' (2), 'Arctic' (1), 'Ecosystem Services...' (1), 'Energy Infrastructure' (1), 'Permafrost and...' (1), and 'Water Resources' (1). The main content area shows '38 datasets found' and lists four datasets, each with a 'Federal' tag. The datasets are: 1) 'EnviroAtlas - MSPA connectivity with water as missing and 1-pixel edge width for the conterminous United States', 2) 'EnviroAtlas - Cultivated biological nitrogen fixation in agricultural lands by 12-digit HUC in the Conterminous United States, 2006', 3) 'EnviroAtlas - Potentially Restorable Wetlands in the Conterminous United States', and 4) 'EnviroAtlas - Biological nitrogen fixation in natural/semi-natural ecosystems by 12-digit HUC for the Conterminous United States, 2006'. Each dataset entry includes a brief description, a link to the U.S. Environmental Protection Agency, and buttons for 'HTML', 'Esri REST', and 'HTML'.

- Tag

- Different tags are available to narrow down the search
- Some available tags are different states of the United States

- Format

- Users can select the desired format of the dataset to narrow down their search.
- Some of the available formats are HTML, EXCEL, CSV, ZIP, and PDF

- Organization, Publisher, Bureaus
 - Users can select datasets by their sources
 - They are divided into 3 different categories, Organization, Publisher, and Bureaus

Downloading from Dataset

- How to download from the dataset
 - Each data collection has a set of data that can be opened with programs such as CSV formatted data and excel.
 - Data can also be downloaded from the Dataset
 - Each data provides different variations of files that can be downloaded

The screenshot displays the 'EnviroAtlas' dataset page. On the left, there is a 'Contact' section with the email 'EnviroAtlas@epa.gov', 'Share on Social Sites' links for Twitter and Facebook, and a 'Dataset extent' map of North America. The map includes a zoom control and is credited to 'OpenStreetMap contributors' and 'Stamen Design (CC BY 3.0)'. The right side of the page features 'Access & Use Information' with a license notice, and 'Downloads & Resources' which lists three 'Web Resource' links, each with a 'Visit page' button, and an 'Esri REST API Endpoint' with an 'Open With' dropdown and a 'Visit page' button.

- Possible programs they can use to analyze the datasets
 - XLS files
 - Zipped CSV files
 - HTML
 - ArcGIS
 - Data.gov Map preview

Examples of Questions and Visualizations

U.S. Bioenergy Statistics Dataset

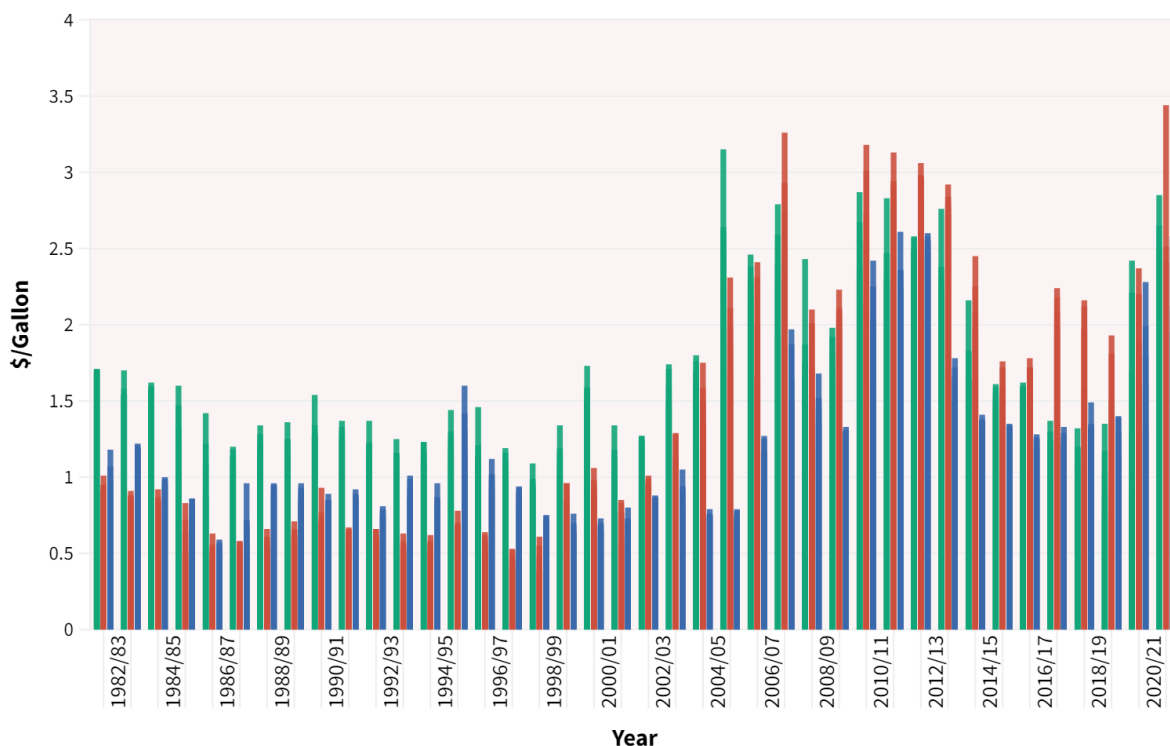
Statistics of different fuel types, their supply and disappearance

- Basics of dataset
 - Entries: Around 2000 entries, 19 tables
 - File formats: Excel

The U.S. Bioenergy Statistics are a source of information on biofuels intended to present a picture of the renewable energy industry and its relationship to agriculture. The statistics highlight the factors that influence the demand for agricultural feedstocks for biofuels production; for instance, numerous tables emphasize the relationship between energy and commodity markets. ERS analysts track U.S. ethanol and biodiesel production, consumption, and trade. They also monitor and analyze U.S. bioenergy policy and events that affect the domestic and international biofuel and feedstock markets.

Ethanol, Gasoline, and Corn Cost per Gallon

■ Ethanol \$ / gal ■ Gasoline \$ / gal ■ Corn \$ / gal



Scientific Question:

When did gasoline prices become more expensive than ethanol?

Analysis:

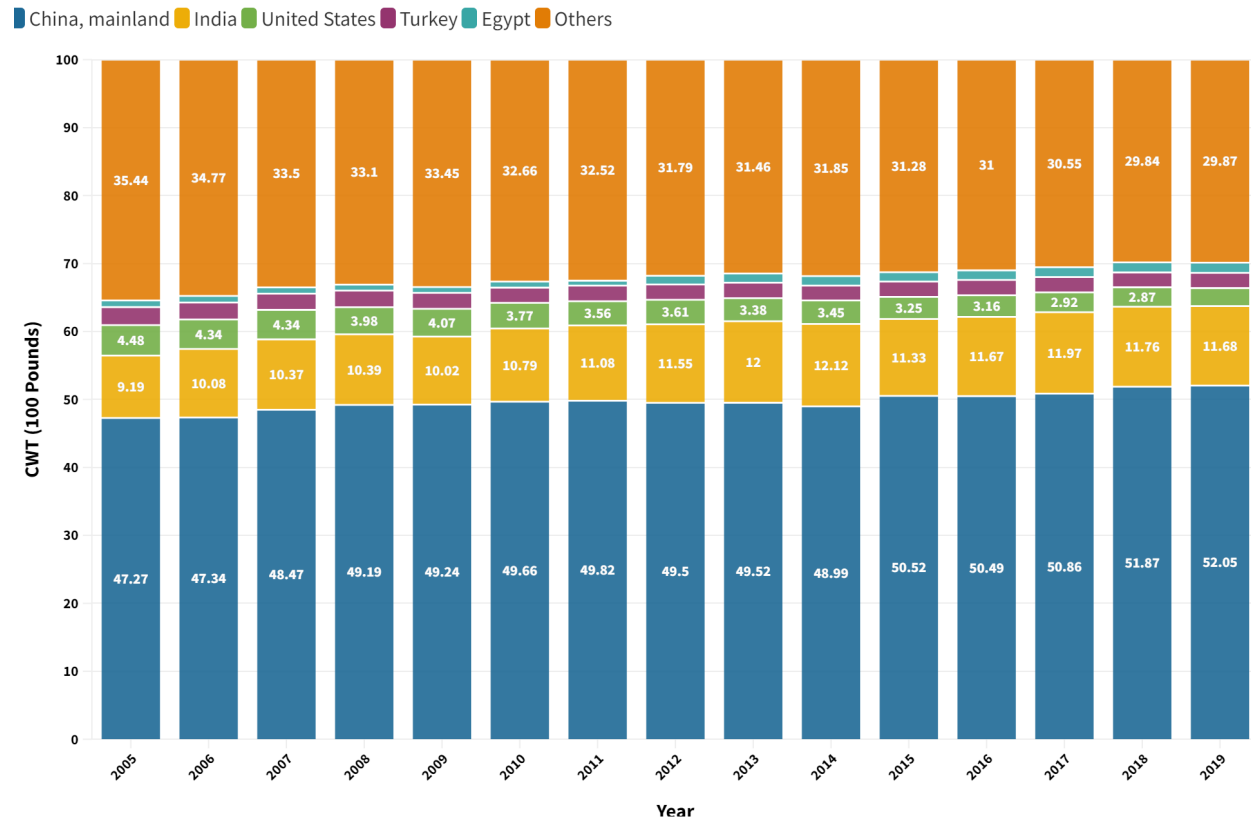
This grouped column chart allows us to compare ethanol, gasoline, and corn fuel values per gallon by year. The graph shows that 2005/2006 was the turning point where gasoline prices were consistently higher than ethanol prices. This group column chart works well along with the colors because it is very clear to see the separate trends while also comparing them to one another.

Vegetable and Pulses Data

Vegetable production throughout the world, such as per capita availability.

- Basics of dataset
 - Number of entries: 5000+
 - File formats: Excel
- The data contains a time series of annual per capita supply and use data for fresh and processed vegetables and for dry pulse crops. Included are U.S. production, exports, imports, per capita utilization, and prices. This data product provides users with comprehensive statistics on fresh and processed vegetables and pulses in the United States, as well as global production and trade data for these sectors. It harmonizes and integrates data from the ERS market outlook program with data collected by different Federal and international statistical agencies to facilitate analyses of economic performance over time, and across domestic and foreign markets. This data was collected by the Economic Research Service.

Vegetable Production by Leading Country



Scientific Question:

Is the United States competitive with other countries in vegetable production?

Analysis:

This stacked bar graph shows the percentage of total vegetable production each country produces by year. From the graph, we can see that China now has more than half the production of the vegetables in the world, and this percentage is steadily increasing. Since 2005, the US has been declining in Vegetable production, compared to every other leading country, which are all increasing production. This may mean the United States is becoming more dependent on other countries for vegetables, since other countries are slowly taking over a higher percentage of produce.

Dataset 3

Atlas of Rural and Small-Town America

- Number of entries: Approx. 3500 N
- File formats: CSV, XLS
- This dataset has statistics about 5 broad categories of socioeconomic factors: People, Jobs, County Classifications, Income, and Veteran status. Each factor has different statistics such as median income and per capita income for the income factor, data on veterans including education and unemployment, and employment trends for jobs in the small towns. This data was collected by the American Community Survey, and the Bureau of Labor Statistics. Using the variables that are given, a user can use the data to look into specific states and counties, and use the median income and poverty rate to see a trend in different categorical variables, such migration and immigration against the per capita income. Also, the ability to zoom into and out of states and regions helps create different stories, such as looking at bigger states against smaller states, or regions in more rural states against regions in less rural states.
- Analysis of columns

Median Income, Per Capita Income, and Poverty Rate By State In Rural Areas



If we analyze the graph above, we can see higher median incomes and per Capita incomes for states that statistically are larger and have a higher poverty line overall; it also correlates to the poverty rates in the states where the median income is lower. If we zoom in more, we can look at smaller counties and notice that the per capita income, especially in smaller states, are lower than the average across America.

- Key scientific insights
 - Is there a significant correlation that can be seen looking at country classifications, and employment and unemployment trends, especially considering the persistent poverty and per capita income?
- The data limitations for this dataset is creating good visualizations that can show good analysis because some of the variables do not always have a direct relationship with the big factors, so to create a good connection, sometimes you need multiple visualizations, around 2-3, to do proper analysis and find a good correlation in the data, even if the correlation can be seen clearly in the data.

FAQ section

1. Where did we find this data collection?
 - a. The collection is located on Data.gov using the filter, "climate".
2. What is data.gov?
 - a. Launched in 2009, data.gov is an open source collection of datasets hosted by the U.S. General Services. From the OPEN Government Data Act, it is required that government database made available to the public
3. Why did we choose this collection (climate)
 - a. Climate change is an important issue that can always be improved with organization and documentation to make it easier and accessible for everyone to read. We felt that the significance of climate change is a time sensitive issue that needs to be addressed quickly.
4. Why are there datasets unrelated to climate change here?
 - a. Climate change is a very broad topic and directly affects a number of different things, such as crops for farming. There may be datasets related to the direct effects of climate change because they are still at its core studies that are done with the effects of climate change in mind and taken account for
5. What is html and why are most datasets in this format?
 - a. HTML stands for HyperText Markup Language, and its primarily used in web development to make websites. It can also be used to store data, such as the datasets found here. Using a regular data program such as Excel, you can easily open, use, and convert HTML files to your preferred file such as .csv or .xls
6. How can we use this data?
 - a. Each dataset in this collection can be used in different ways. Some have direct API's that you can use and import directly when creating a mobile or desktop application, and others can be downloaded directly as a .xls, .csv, or other file type for use in data wrangling programs such as tableau.

Version history

Version	Year
1.0	2022