GHG Agriculture Manual

Throughout this notebook, we have written down notes on our methodology and how to use the resources we’ve obtained. We’ve put “possible work on” and “work on” as todos for feature students who are tasked with this project.

# Folders in ca\_agriculture

## code\_results

This is where the final results should go in. For example, you will find the agriculture\_activity\_emissions and can find the livestock emissions per county by millions of tonnes of CO2e.

## data

We put datasheets that we used to filter/categorize e.g. the Nitrogen Table where we created the crop categories for nitrogen fertilizers to get their fertilizer average.

### Uploaded Data

This folder contains RAW CSVs that are needed to run the Colab Notebook. For now, as of updating this manual, it should only contain:

* Agriculture Census Data (with label).csv
* GHG Emissions (2000-2021).csv
* Nitrogen Application.csv

## Cool Climate GHG Inventory.ipynb

This is where the bulk of the work of joining tables, filtering, and getting the final output results was conducted. Not all work was done in this notebook though as we did some of the work in the sheets of data e.g. averaging fertilizer rates for categories of crops in the Nitrogen Table.

# Using the Colab Notebook

The Colab notebook was created mostly to help simplify joining tables.

## Before Use

There is a folder in the data folder called ***Uploaded Data***. Every time before you start running the code for the notebook, please upload the CSVs in that folder into the “..“ folder so that the data is uploaded.

## Possible Work On - API Implementation

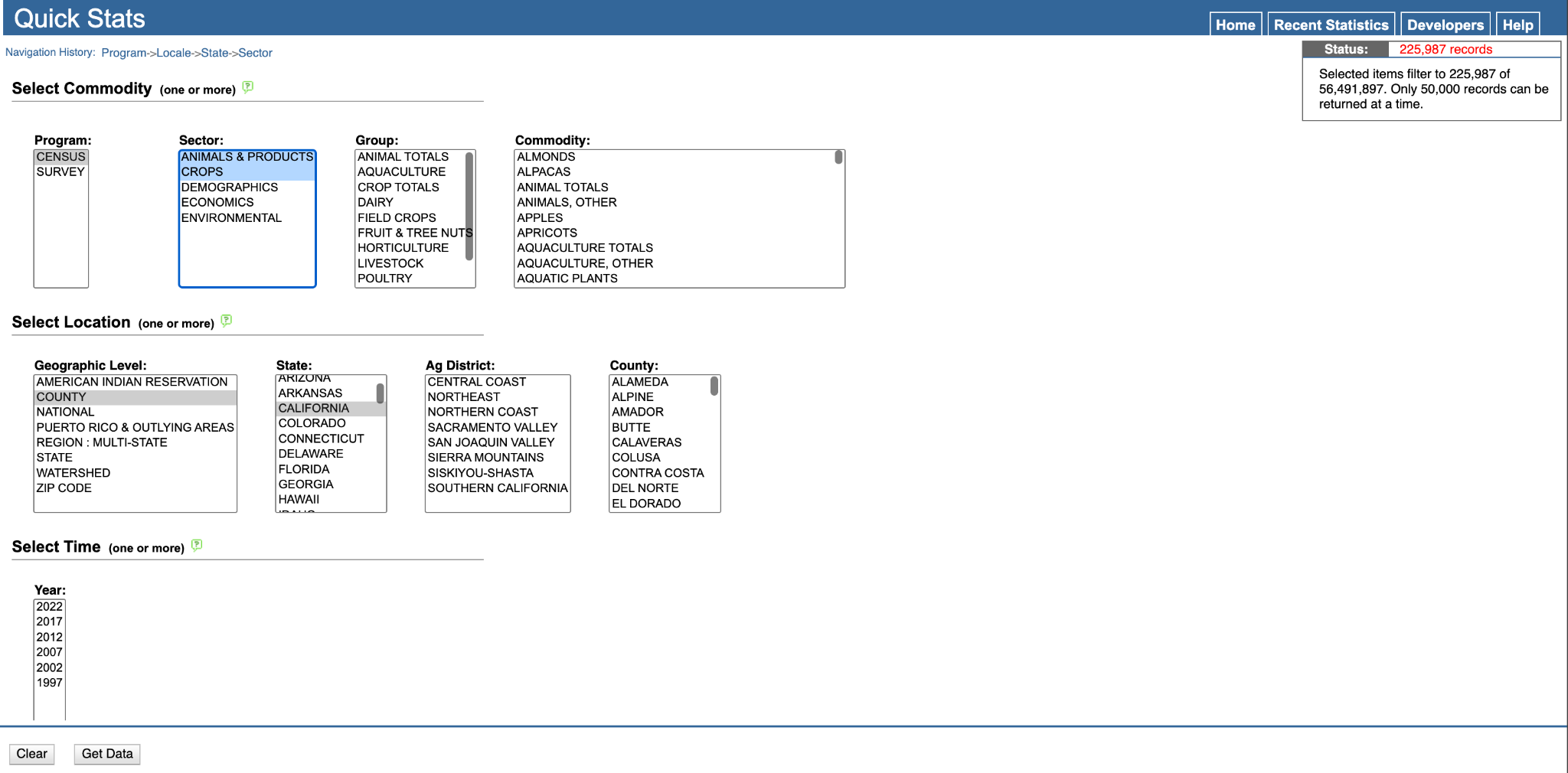
We tried implementing a Quickstats API for the program but unfortunately ran out of time for the time we had so for future students working on this, please consider implementing the [USDA Quickstats API](https://quickstats.nass.usda.gov/api) so you don’t have to upload the agriculture CSV file every time to use the notebook.

# Data

## Datasets

### USDA Quickstats Agriculture Data

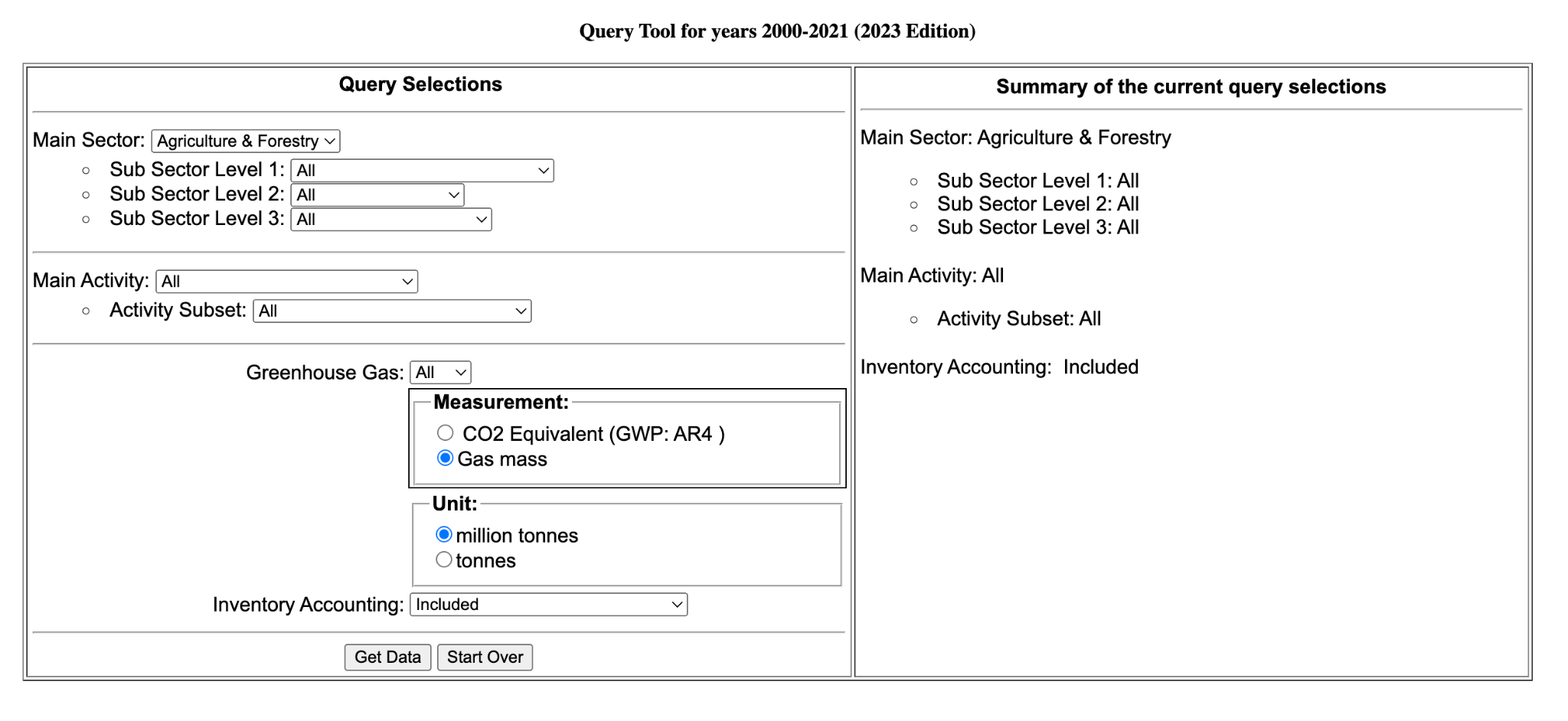
Here’s the link to the [dashboard](https://quickstats.nass.usda.gov/). For this data we did a census and not a survey because while the survey has data for every year, it only does a survey for a few select random counties each year while the census tracks every county every 5 years. The filter type you use to get the data should be below.



As you can see, Quickstats doesn’t allow you to download that many records at once so what we did is we just downloaded multiple counties separately and combined it in the end into one file.

### California Air Resources Board GHG Emissions

For this data, we used this [link](https://ww2.arb.ca.gov/applications/greenhouse-gas-emission-inventory-0). We were also provided a [link](https://ww2.arb.ca.gov/applications/greenhouse-gas-emission-inventory) to data from 1990 but did not use this in the actual project. To get the data, please make sure you select “Agriculture & Forestry” for the main sector and “Gas Mass” for measurement. We convert the gas mass into CO2e manually according to AR5 GWP.



## Livestock Emissions

### Methodology

Create a pivot table by joining the filtered agricultural census data from USDA Quickstats only to include livestock and combining it with the emissions data for non-crop/others from the California Air Resources Board.

**For counties that did not disclose their data (D) or had less than half the rounding unit (Z), we just input values of 0. If you would like to change this methodology, please do so.**

### Livestock Labels

* Cattle
* Poultry
* Swine
* Other Livestock
  + This refers to equines, goats, and sheep
* Other Livestock\*
  + Anything that doesn’t identify with the above
  + We didn’t use this label for the final code because of complexity and also the emissions data not having this.

## Crop Emissions

### Methodology

For crops, we associated emissions from crops due to fertilizer usage.

Acquired fertilizer emissions from various sources, prioritizing California-based fertilizer emissions. The average fertilizer emission rate was used for calculation. Sources are linked in the table. We also added this into the Agriculture Census Data (with sublabel) xlsx fore easier reference but please make sure to check which one is the more updated one (should be the census data)

**For counties that did not disclose their data (D) or had less than half the rounding unit (Z), we just input values of 0. If you would like to change this methodology, please do so.**

### Sublabels

Categorized according to this website <https://alliancebioversityciat.org/stories/how-many-types-vegetables-are-there-complete-guide> and other methods

Final Sublabels

* Berries
* Bulb Vegetable
* Citrus
* Drupe
* Flower Vegetable
* Fruit Vegetable
* Grain
* Herbs
* Leafy Greens
* Legumes
* Melons
* Pome
* Root Vegetable
* Seed Vegetable
* Stem Vegetable
* Tropical
* Field Crops
  + This was not used in the full fertilizer data but implemented in the full agriculture census data since there was a field crop total field.
  + Field crops average are just the average of the Grain, Legumes, Root Vegetable, and Seed Vegetable averages.

California Agricultural GHG Emissions

Customer class: All customers in the Socal Gas service territory. Data Range: 1/1/2005 – 12/31/2023

Data Points: County, City, Year, Sector, Commodity (Gas or Electric), Count of Customers by Energy Provider

(PG&E or CCA or Other), Total Usage (Therms or kWh), Average Usage (Therms or kWh), CCA Code to CCA name cross-reference, SoCal Gas CO2 Emissions Factor, by year

Total annual consumption (Therms) by sector (residential, industrial, commercial, agricultural), and count of Customers.

## Filtering the Crops

Many of the crops have multiple fields. For example, berries like strawberries could have acres grown, acres harvested, and acres non-harvested. For the context of this project, we looked at only fields that had the text of unit “acres” in its Data Item column because the Nitrogen Table is in (lbs/acre). Hence, we first filtered down to get only crop fields with acre in them.

### Possible Work On - What Acre label to use

Given the complexity and limitation of time, we decided to use the acre label with the max yield through a groupby. In an ideal situation, we would go through every label and use ones that made sense e.g. using “acres bearing and non-bearing” instead of “acres bearing” or “acres non-bearing” for fruits because “acres bearing and non-bearing” is the sum of those two. However, we simply didn’t have enough time so we just maxed it. Please if you can help improve this methodology in the Crops Section of the Code.

### Work On - Liming

In the California Air Resources Board emissions data, there are two fields in there referred to as Liming. Liming is not the same as nitrogen fertilizing; however, both are applied to soils for crops. **We made a very large assumption that, for example, if a county uses a significant proportion of nitrogen fertilizer, then they use a significant amount of liming**. This is definitely not the best method; however, given the time constraints, we opted for this path. For future purposes, please definitely change this. One idea of a methodology that we think might work is to identify crops that have a heavy presence of liming, filter them down, and then do the same process listed above that we did for nitrogen fertilizer acres.

### Crop Residue

We assumed that based on the proportions of crops if a county had more crops, there would be more crop residue. Hence we grouped this with crop fertilizer usage. Feel free to change the methodology on this one.

### Work On - Histosols

We did not have the expertise in how to deal with histosol emissions from the GHG emissions data so we didn’t do it for now. Definitely work on this.

### Rice Emissions

We used WILD RICE and RICE to calculate our rice emissions from the GHG emissions data. Similar methodology to the fertilizer data. Some counties had gaps of zeros in the pivot table because they didn’t disclose data for that year even though we suspect they still had some rice production. We, however, didn’t do anything to replace those zero values because 1. Majority of rice production is concentrated in Sacramento Valley counties so it wouldn’t be too much of a difference 2. We ran out of time.

walnut\_production = agriculture\_cleaned2[(agriculture\_cleaned2['Commodity'] == 'WALNUTS')]

walnut\_production = walnut\_production.groupby(['county', 'year'], as\_index=False)['yield'].agg(sum)

walnut\_production = walnut\_production.pivot(columns = 'year', index = 'county', values = 'yield').fillna(0)

walnut\_production = walnut\_production.div(walnut\_production.sum(axis=0), axis=1)

walnut\_burning = pd.DataFrame()

for col in walnut\_production.columns:

if (col != 1997):

walnut\_burning[col] = walnut\_production[col] \* inventory\_2000\_2021.loc['Crop Burning - Walnut', str(col)]

for county in county\_list:

if county not in walnut\_burning.index:

walnut\_burning.loc[county] = [0, 0, 0, 0, 0]