Crop Categorization

To simplify our analysis given the large number of crops in the landIQ, we categorize into 23 broader crop categories.

Economic dataset

The USDA dataset provides crop data at a county level. Due to the extensive variety of crops, we aim to select a representative, or "proxy," crop for each of the 21 defined crop categories on a county level. This process involves first grouping USDA crops into these 21 categories. For each category, we then identify a proxy crop that best represents the overall characteristics of the group. This is achieved by calculating the acreage-weighted price yield for each crop category and selecting the crop with the closest match to this value.

[Eq. 1]

* Should we pick proxy crops on a county level or hydrologic region?

Spencer said he chose 1 proxy crop for all central California in an email. In the document he is doing for each hydrologic region

1 proxy crop for all central California may not be the best representation for categories that have multiple crops (ie orchards, cucurbits, subtropical).

If we pick one proxy crop all throughout central California, crop variation can be high between hydrologic regions. For example, in Sacramento, plums make up a significant percentage of orchard crops, a trend that doesn’t hold in other areas such as San Joaquin or Tulare counties. Moreover, many counties in San Joaquin and Tulare may not have price and yield and thus we will have mask them with average values from other counties.

If we pick one proxy crop at the hydrologic region, crop variation can be high between counties. Take plums, for example—they make up a big part of orchard crops in Sacramento County, but in other counties in the same region, there might not be any plums at all. And we will have fill it by averaging values from other counties.

This could be solved if we pick a proxy crop on county level.

* Why calculate acreage weighted yield for each crop
  + Spencer has done this but doesn’t provide an explanation for where we’re using it.
* Spencer is averaging for 2016-2019 data, we’re not doing that right?
* Why tomatoes and grapes are treated differently.
  + Spencer: Tomatoes and grapes were noted as two commodity types which have notable differences in economic value given their end-use purpose (e.g. fresh vs. processing tomatoes or table vs. wine grapes). Both commodities are reported only generally in the LandIQ dataset. To remedy this discrepancy, harvested acreage from USDA NASS for counties within each hydrologic region were summarized and assessed to determine approximate percentages grown for fresh market versus processing and table versus wine for tomatoes and grapes, respectively
* Irrigated pasture is not included as a commodity in USDA NASS statistics.
  + Spencer: prices and yields were instead estimated from county agricultural commissioner reports as $215/ton and 3.5 ton/ac, respectively

Land Cover

LandIQ parcel scale land cover for 2020 was intersected with county boundaries using ArcGIS Pro.

If we work on identifying proxy levels on hydrologic region scale, then each county is assigned a hydrologic region based on the region with the largest spatial overlap. Please find the image below.

A map of the state of california

Description automatically generated

Parcels with missing economic data

In counties with parcels lacking economic data, we use the following sequence to fill gaps. If one step yields NaN, we proceed to the next:

1. Use the average economic data of crops from intersecting counties.
2. Apply the average economic data of crops within the parcel’s hydrologic region.
3. Use the average economic data of crops from intersecting hydrologic regions.
4. Apply the average economic data of crops across all hydrologic regions.

~~Parcel Scale~~

~~We set up a 50-mile radius around parcels that don’t have economic data and average the data for parcel that have it withing the range.~~

~~A map of the state of california

Description automatically generated~~

~~County scale is easier to implement. I don’t think there will be difference in accuracy.~~