Prompt 1

Author -- Elaine Swanson
HDSI Agri Datathon
October 2024

Background Info:

VegScape and CropScape are two USDA NASS web geospatial processing services that greatly facilitate remote sensing data accessing and processing, product data querying, visualizing, and disseminating. To learn more, read the papers about the VegScape and the CropScape web services for all functionalities. Together, the satellite image data pulled from these two web services can be used to produce crop phenology insights. Crop phenology is the study of the physiological stages of crop growth and development from planting, vegetative development, flowering, to harvest. It's a vital part of crop growth management and yield estimation, as it can be integrated with information about how environmental factors and management practices impact plant performance. See the panel diagram below to visualize the crop phenology for a field crop of potatoes (provided by Llama 3.1).

Image 1: Crop phenology images.

Prompt 1

Author -- Elaine Swanson
HDSI Agri Datathon
October 2024

When satellites observe and record terrain that contains cropland and natural vegetation, researchers can measure the development of plant life, or "greenness", via an index called the normalized difference vegetation index (NDVI). Please read this [Wikipedia] article to understand more about the NDVI index, but this is the measurement that is already calculated for your team via VegScape. Researchers use this index to create time series graphs mapping crop phenology, we will refer to these as "crop phenology reports". They use these to make comparisons from year to year. Such as the graph below. **For example:** one can see a 0.2 index difference between the year 2010 and the year 2023 in the month of March. Knowing this, we can then investigate some questions and gain insights. **For example:** Does this result in differences in final crop yield or crop health during these years? Is this a result of a change in winter season's snowfall?

Overall, crop phenology reports show the general trend of agricultural plant growth from satellite data; rising levels in the spring, peaking in the summer, and harvesting in the fall (drops off rapidly from Nov-Dec).

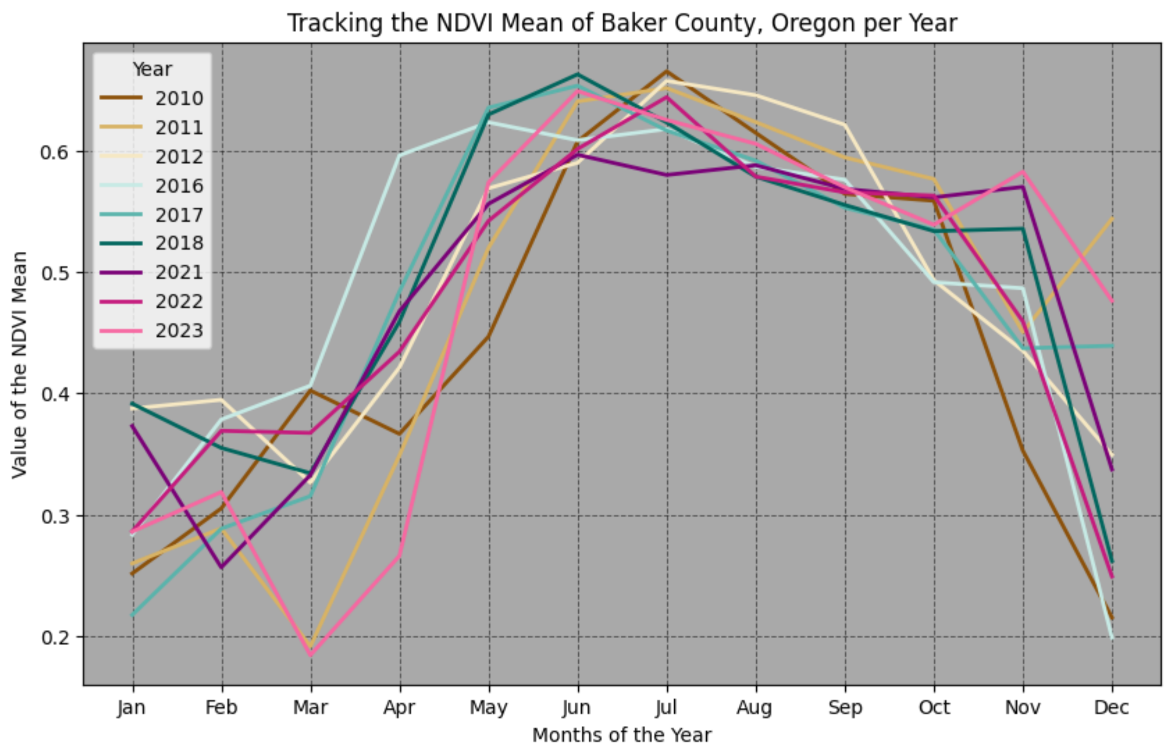

Image 2: crop phenology report for Baker County, OR 2010-2023

Prompt 1

Author -- Elaine Swanson
HDSI Agri Datathon
October 2024

Currently, these two web services are not integrated, and there is no current capability to distinguish **crop NDVI values** from those **NDVI values of natural vegetation**. Your task is to create a scalable or parallel framework in Python or C++ that can integrate these two datasets over 23 counties from the Western United States, more details follow. These counties are located in the Western states of Washington, Oregon, and Idaho and can be located in the map below (provided by mapchart.net).

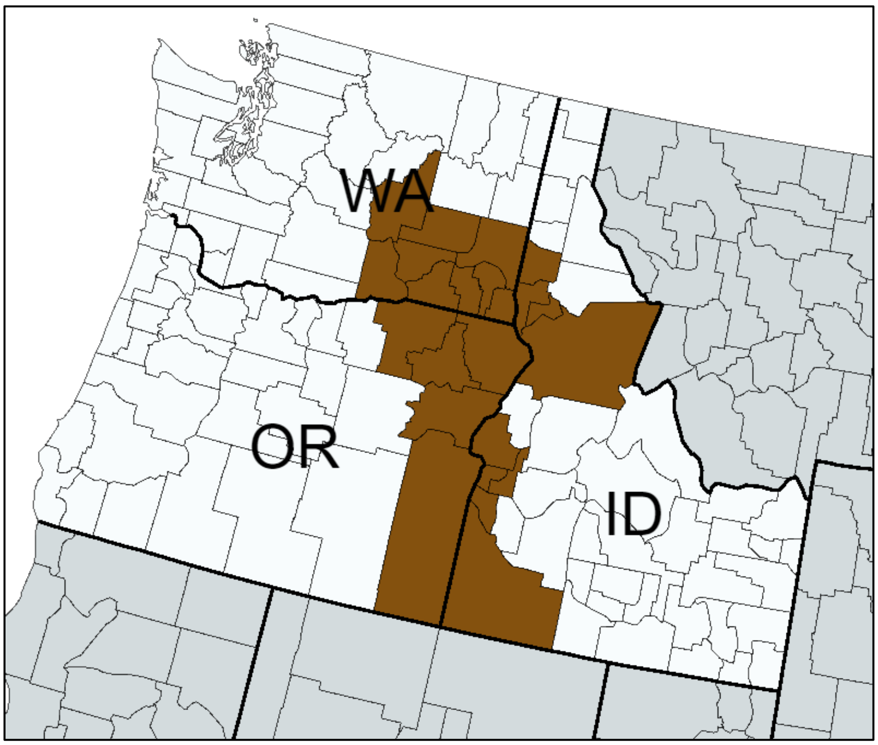

Image 3: Location of counties of interest.

Author -- Elaine Swanson
HDSI Agri Datathon
October 2024

Washington State	Oregon State	Idaho State
Grant	Baker	Latah
Adams	Malheur	Nez Perce
Benton	Umatilla	Lewis
Franklin	Union	Idaho County
Whitman	Wallowa	Washington County
Walla Walla		Payette
Columbia		Gem
Garfield		Canyon
Asotin		Owyhee

Table 1: List of counties of interest.

These counties were chosen because they share a climate that is shaped by their position in the rain shadow of the Cascades Mountain Range, resulting in dry conditions, hot summers, and cold winters. The terrain is defined by rolling hills and valleys, and important river systems (Columbia and Snake Rivers) that enable irrigation-based, monoculture agriculture in an otherwise dry region.

How to Complete this Prompt:

Your team will develop a pipeline that executes the following steps:

- Isolate the agricultural activity of the CropScape images based on the land classifications. You do not have to be so granular and separate by crop type. Just a binary, what is active agriculture and what is not.
- Extract the NDVI values from similar areas from the VegScape images.
- Create crop phenology time series (max, mean, min) NDVI dataframe reports per county over all years. Proceed to use a model of your choice to analyze past trends and predict last year's (2023) crop phenology NDVI mean for all counties. Your team

can visualize your 2023 prediction from one county, or a grouping of counties by state, of your choice to the actual 2023 county data.

- The competition won't judge computational efficiency based only on whether your method uses parallelism. There are many scalable methods that don't use explicit parallelism but still handle large problems effectively. Think creatively about how your methods can scale up to handle larger problems.
- You may require significant computational resources, such as GPUs or a large number of CPUs, to effectively process and integrate the **full** dataset.
- The purpose of this exercise is to gain an understanding of the importance of crop phenology to assess food security in the American West, as well as being aware of an important vegetation index derived from satellite imaging.

What your team has:

Cropland Images:

Inside the prompt_1 data folder, you will find the raster cropland images in GeoTIFF format from CropScape for each of the 23 counties over 10 or 11 years. These years are: 2006, 2007, 2010, 2011, 2012, 2016, 2017, 2018, 2021, 2022, and 2023. In total 177 images. Each image represents the crop layout and classification of one county per year, as croplands in this region are typically fixed for the calendar year. Each image is formatted as a standard raster image called a GeoTIFF and has a .tif file extension. **Please note the spatial resolution of these images.** See example images below for six years for Baker County, OR, along with its associated colorbar. You will find that each image is accompanied with a .tif.vat.dbf file. This file contains the crop classification and their associated values of the land surface. Some examples of what you will see in this file are in the table below.

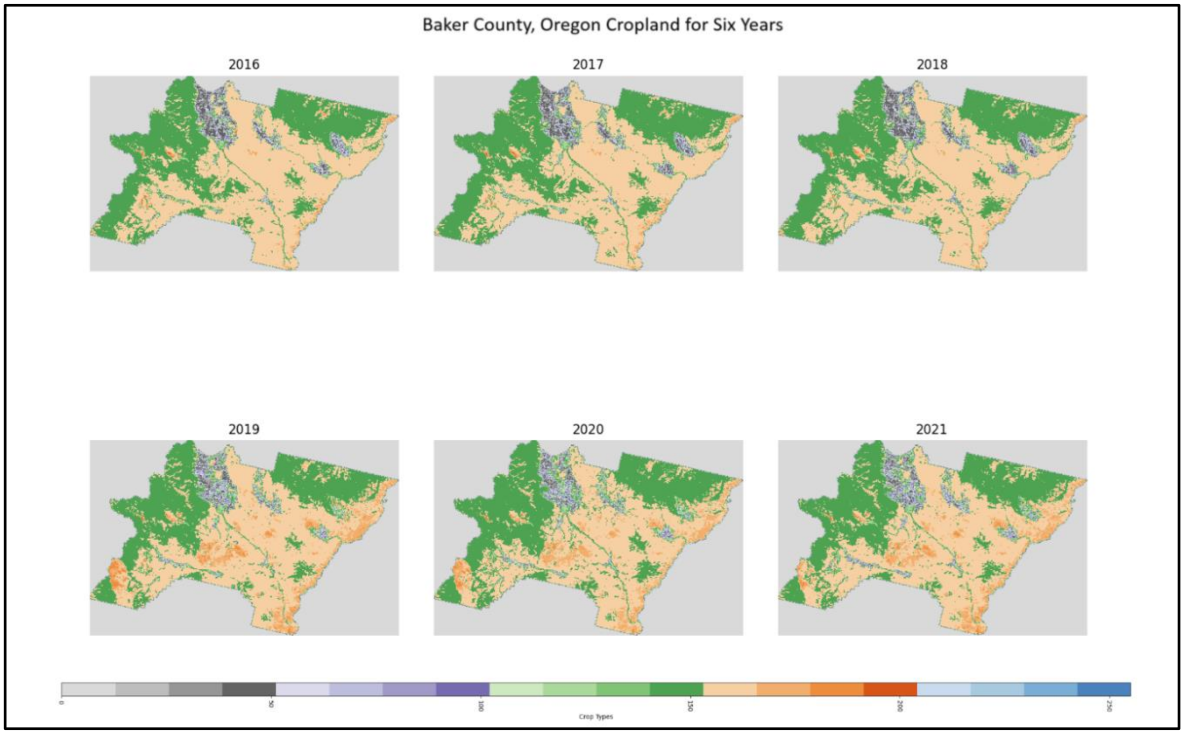

Image 4: Cropland examples.

Pixel Value	Red Color	Green Color	Blue Color	Crop Classification
22	0.537	0.384	0.329	Durum Wheat
43	0.439	0.149	0.000	Potatoes
142	0.576	0.800	0.576	Evergreen Forest
152	0.776	0.839	0.620	Shrubland

Table 2: Land classification examples.

Author -- Elaine Swanson
HDSI Agri Datathon
October 2024

NDVI Images:

Also, inside the prompt_1 there are around 52 NDVI images per year, per county from VegScape for each of the 23 counties over 13 years. These years are: 2001, 2005, 2006, 2007, 2010, 2011, 2012, 2016, 2017, 2018, 2021, 2022, 2023. In total, around 14,476 images. Examples of NDVI images are below for Baker County, OR, along with its associated color bar. **Please note the spatial resolution of these images.** Also note the scale of the NDVI values on the color bar and the values that NDVI can take on the Wikipedia page (the scale of the y-axis on the crop phenology report above). You must make the scale be an accurate portrayal existing in the interval 0 to 1 to produce the crop phenology time series graphs. This document should provide you with more info to scale appropriately *(hint: slide 5)*.

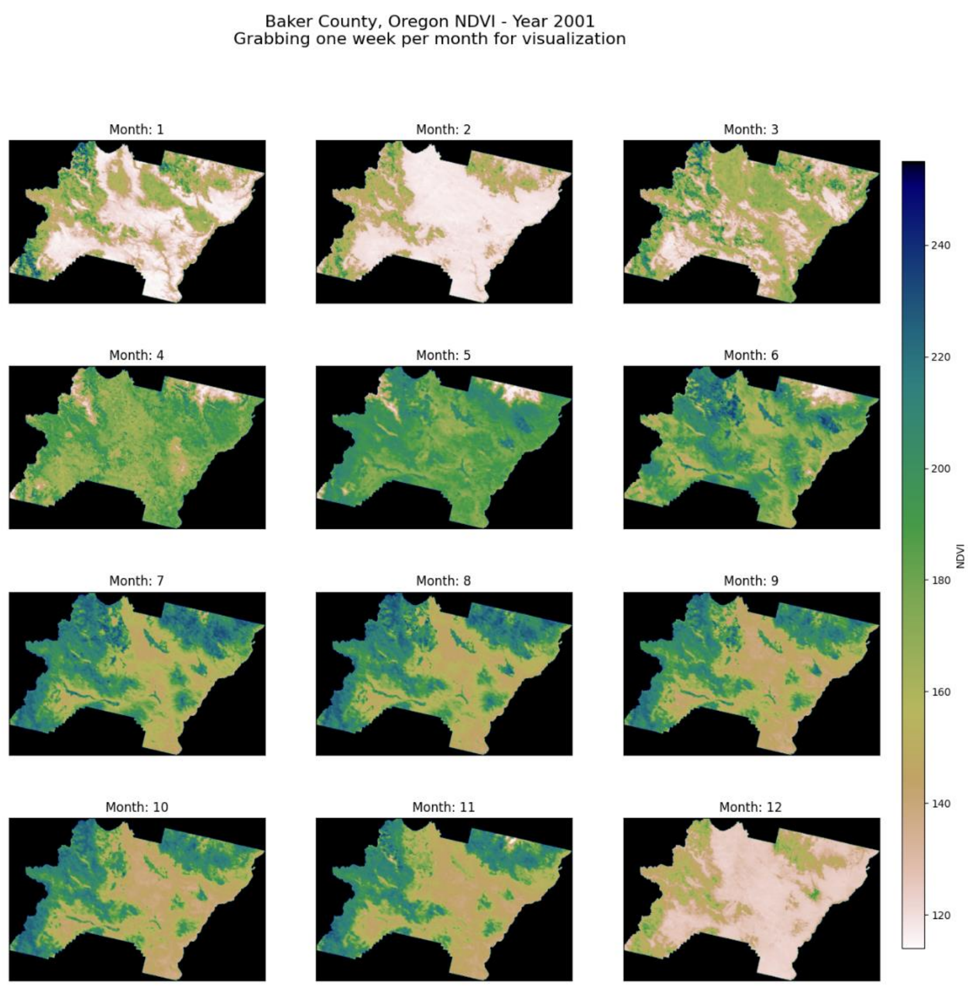

Image 5: NDVI examples.

Prompt 1

Author -- Elaine Swanson
HDSI Agri Datathon
October 2024

Notes:

- There are not cropland image years for every NDVI image year. You may make assumptions on the crop layout for missing years.

Viz Tips:

- The Python matplotlib colormaps used in this document for the cropland and NDVI images were `cmap='tab20c_r'` and `cmap=plt.cm.gist_earth_r` respectively, if you wish to recreate similar images.
- To create your team's final visualizations, it will be beneficial to map your values across the counties of the contiguous United States. **A choropleth map is a type of map that uses color to represent data across geographic areas. The term comes from the Greek words choros (region) and plethos (multitude).** Using a tool like Python Plotly's choropleth maps can enhance your visual analysis. For more information and examples, visit Plotly's choropleth map documentation and ways to create USA County Choropleth Maps in Python.