

! Important: Data Loading and Formats

This section provides crucial information about the available data formats and how to load the Parquet files in R. Please review this carefully before proceeding with your analysis.

Info About Dataset Format

This dataset is available in two formats:

- **Multiband GeoTIFFs:** The original data is stored as multiband GeoTIFF files, with each file representing a specific week's average from February 18th, 2000 to December 31st, 2023. These files were pulled directly from Google Earth Engine. The total size of the GeoTIFF dataset is 11.7 GB, and is split across 1244 files representing each week in date range.
- **Parquet Files:** For easier analysis, the data has been preprocessed and re-exported into annual Parquet files. Each Parquet file contains a full calendar year of data (e.g., `illinois_2000.parquet`, `illinois_2001.parquet`, etc.). The total size of all 24 Parquet files is 12 GB.

Important Note: During the Parquet conversion, all pixels with no data across all 16 bands were omitted. This primarily filters out areas outside the Illinois boundary but within the GeoTIFF's bounding box. Therefore, the Parquet files contain only pixels with at least one valid data point across all bands.

Loading Parquet Data in R

To read the Parquet files in R, you'll need the `arrow` library. Here's how to install and use it:

```
# Install the arrow library (if you haven't already)
install.packages("arrow")

# Load the arrow library
library(arrow)

# Read the Parquet file
df <- read_parquet("data/transformed_data/illinois_2003.parquet")

# View the size of the dataset
```

```
print(dim(df)) # Prints the number of rows and columns

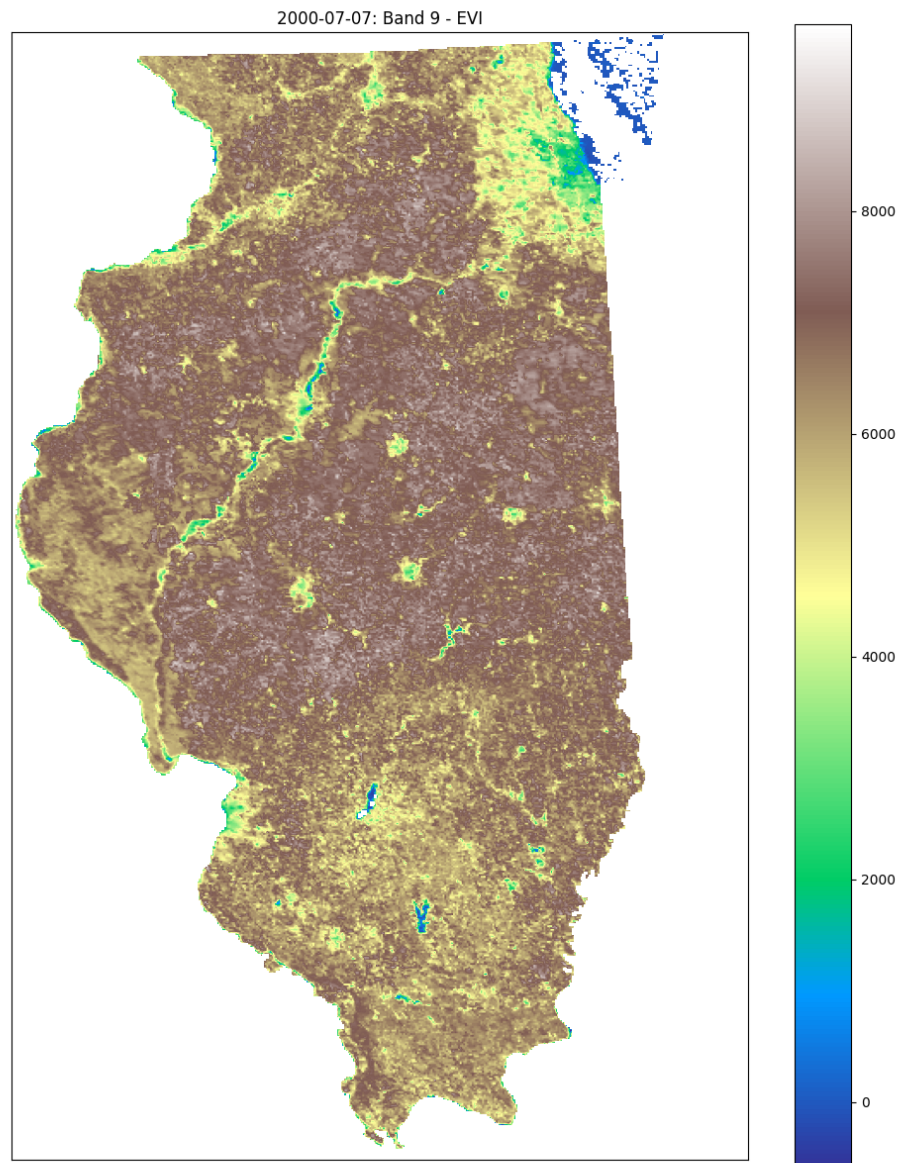
# View the column titles
print(names(df))

# View the data (in RStudio)
View(df)
```

i Note

I've tested this on my machine and was successfully able to load the converted Parquet files into RStudio using the `arrow` library.

Data Dictionary for Illinois GeoTIFF Metadata



General Metadata Summary

- **Geographic Coverage:** Illinois, USA
- **Coordinate System:** EPSG:3347 (Canada Lambert Conformal Conic)
- **Spatial Resolution:** 1 km x 1 km
- **Temporal Coverage:** February 18th, 2000 to December 31st, 2023
- **Temporal Resolution:** Weekly
- **Data Sources:**
 - Daymet v4 (daily meteorology, averaged to weekly)
 - MODIS (vegetation indices, ET, FPAR, LAI; composites downscaled spatially, up-scaled temporally)
 - GFSAD (2011 land cover, reclassified)
- **Rendered GeoTIFF Dimensions:**
 - Width: 431 pixels
 - Height: 660 pixels
 - Bands: 16
- **NoData Value:** None Used (NaN when loaded)

Band Explanation

Band Name	Description	Units	Data Type	NoData Value
crop	Land cover type (1: non-cropland, 2: irrigated cropland, 3: rainfed cropland)	Category	float32	None
dayl	Daylength	seconds	float32	None
prep	Precipitation	mm/day	float32	None
srad	Downward shortwave radiation at the surface	W/m ²	float32	None
swe	Snow water equivalent	kg/m ²	float32	None
tmax	Daily maximum 2-meter air temperature	degrees C	float32	None
tmin	Daily minimum 2-meter air temperature	degrees C	float32	None
vp	Water vapor pressure	Pa	float32	None
EVI	Enhanced Vegetation Index	unitless	float32	None
NDVI	Normalized Difference Vegetation Index	unitless	float32	None
ET	Evapotranspiration	mm/day	float32	None
LE	Latent heat flux	W/m ²	float32	None
PET	Potential evapotranspiration	mm/day	float32	None
PLE	Potential latent heat flux	W/m ²	float32	None
FPAR	Fraction of Photosynthetically Active Radiation	unitless	float32	None
LAI	Leaf Area Index	m ² /m ²	float32	None

Band			Data	NoData
Name	Description	Units	Type	Value

Data Source and Processing Summary

- **Data Acquisition:** All data was obtained from Google Earth Engine using their Python API.
- **Spatial Extent:** The dataset covers the entire state of Illinois, with the boundary defined by the ‘TIGER/2018/States’ census dataset.
- **Temporal Extent:** The dataset spans from February 18th, 2000 to December 31st, 2023. It is organized into weekly averages, resulting in 1244 datasets (originally 1245, but week 71 was removed due to missing FPAR data).
- **Temporal Aggregation:** Daymet data, originally daily, was averaged to weekly values. MODIS composite data was downsampled.
- **Spatial Resampling:** MODIS ET and FPAR indices, originally at 500m resolution, were downsampled to 1km resolution using bilinear interpolation to match the resolution of other datasets.
- **Land Cover Data:** Land cover data is from the 2011 NASA GFSAD product. It was reclassified into three categories: non-cropland, irrigated cropland, and rainfed cropland.

Notes on Original Datasets

1. Daymet v4 (Meteorologic Data)

- North America
- 1980.01.01 to 2023.12.31
- 1000m x 1000m, Daily
- EEID: [NASA/ORN/DAYMETS_V4](#)

2. MODIS (Terra 1km) - EVI/NDVI

- Global
- 2000.02.18 to 2024.10.31
- 1000m x 1000m, 16-day Composite
- EEID: [MODIS/061/MOD13A2](#)

3. MODIS (Terra 500m) - ET/LE

- Global
- 2000.01.01 to 2023.12.27
- 500m x 500m, 8-Day composite
- EEID: [MODIS/061/MOD16A2GF](#)

4. MODIS (Terra 500m) - Leaf Area and FPAR

- Global
- 2000.02.18 to 2024.11.08
- 500m x 500m, 8-Day composite
- EEID: [MODIS/061/MOD15A2H](#)

5. Global Food-Support Analysis Data (GFSAD) Cropland

- Global
- 2010.01.01
- 1000m x 1000m, one-time (2010)
- EEID: [USGS/GFSAD1000_V1](#)

Types of Vegetation Metrics

- **Evapotranspiration (ET):** Measures the total water loss from soil and plants; higher ET indicates active plant growth and sufficient water availability. Over 25 years, an overall decrease in ET due to climate change could indicate reduced plant activity or increased water stress.
- **Fraction of Photosynthetically Active Radiation (FPAR):** Represents the fraction of sunlight absorbed by vegetation for photosynthesis; higher FPAR denotes healthier, more productive crops. Over 25 years, a declining FPAR trend could indicate an overall reduction in crop canopy density.
- **Leaf Area Index (LAI):** Quantifies the total leaf area per unit ground area; higher LAI reflects denser foliage and robust plant growth. Over 25 years, an overall decrease in LAI over time could point to diminished vegetation cover or stunted growth.
- **Enhanced Vegetation Index (EVI):** An index that enhances vegetation signals by reducing atmospheric and soil background noise; higher EVI values correspond to healthier vegetation. Over 25 years, a downward trend in EVI might indicate declining vegetation vigor.
- **Normalized Difference Vegetation Index (NDVI):** Assesses vegetation greenness by comparing red and near-infrared reflectance; higher NDVI values signify healthier, greener vegetation. Over 25 years, a decrease in NDVI over 25 years could suggest reduced vegetation health or coverage.

NDVI and EVI capture vegetation greenness and vigor, offering a direct assessment of plant health, whereas **LAI**, **FPAR**, and **ET** provide information on structural and functional aspects of vegetation. The scientific basis for preferring NDVI/EVI lies in their sensitivity to chlorophyll content and ability to detect changes in vegetation health more directly. The general hypothesis is: if the climate in 2000 was more favorable for crops than in 2024, we would expect to see higher values of ET, FPAR, LAI, EVI,

and NDVI in 2000, reflecting more vigorous plant growth, denser foliage, and healthier vegetation due to optimal growing conditions.