**Goals:**

* Model county soybean yield as a function of weather, soil, and various remote sensing datasets.
* Compare predictions for different lead times (harvest is generally in late October- November, so you can use data through October, vs. through August, vs. through July or June)
* Compare predictions with and without using the satellite greenness measures

**Rationale:**

Soybean yields are affected by many variables, including weather, soils, and management. These factors interact, so for instance dry weather may be very harmful in some soils but less so in others. Also, there are many potentially influential factors such as pests for which good data do not yet exist, but these “omitted” variables have some spatial coherence. Thus it’s likely that models that do a good job at capturing nonlinearities and interactions, and that account for spatial/temporal structure, will perform better than those that do not.

**Data sources:**

* Soy yield data <https://quickstats.nass.usda.gov/> (use survey, not census, and use yield for rainfed areas)
* 1km gridded soil properties: <http://soilgrids.org/>
* satellite based surface temperature: <https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/myd11c3>
* air temperature and rainfall: <http://www.prism.oregonstate.edu/recent/> or
* modis satellite greenness measures: <https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod13c2_v006>
* land cover data: <http://www.mrlc.gov/nlcd11_data.php> or <http://landcover.usgs.gov/global_climatology.php> (useful for masking gridded datasets to average values over cropland areas)