1 Crop Type and Economic Development

The overall question is whether specific crops (wheat vs. rice) have influenced economic development over time. Some examples of how this might work:

- Gender roles and the plow (Alesina et al): Show that ethnic groups that farmed crops that were compatible with heavy plows have different attitudes towards women than those who do not require heavy plows.

- Labor intensity (Vollrath, 2009; Eberhardt and Vollrath, 2015): The elasticity of agricultural output with respect to labor appears to differ by crop type. Wheat has a low elasticity (so additional labor does little to increase yields) while rice has a high elasticity (additional labor increases yields a lot). This changes the relative cost of agricultural goods in terms of labor, and can influence the speed of development. In places with high elasticity, labor is not released as quickly to non-agriculture and food costs are low so fertility remains high.

- Crops have different requirements for irrigation/water use (Bentzen etal 2014): may influence need for collective action and/or change ability to extract rent from users of water.

- Variation in crop types within areas may influence state formation (Litinia 2014b): variation encourages trade and specialization, but requires more advanced states to manage.

- Urbanization related to agricultural productivity (Florax et al 2014)

1.1 Shape Files

In Data folder (within Dropbox folder), gadm_v2_shp is country-level shape-file, but includes sub-national units for nearly all countries.

Geographic ethnic boundary data is in the GeoEPR folder in Data folder. Should be boundaries of ethnic groups within all countries.

1.2 Economic Outcomes

Source for urban, rural populations by grid-cell from HYDE database:

http://themasites.pbl.nl/tridion/en/themasites/hyde/download/index-2.html

Also includes data on crop area use. Use "Guest" to access to FTP site. At that point, there are various iterations of the HYDE data (concave, constant, etc.) and I have not explored enough to know the meaning of that variation.

Night lights, used by others to map geo-spatial economic activity (rather than at the country level). Site is:

http://ngdc.noaa.gov/eog/download.html

Average visible, stable lights, and cloud free coverages. I believe the F10-F18 refer to the separate satellites.

1.3   Agricultural Production Data

Ramankutty et al (2002) is a standard reference for agricultural suitability. It is an overall measure, based on the potential for vegetative growth, and not crop specific. It also does not depend on patterns of rainfall, only totals. In Data folder, as "suit".

The FAO's GAEZ project does specific crop-based measures for whole world.

http://www.fao.org/nr/gaez/en/

The Data Portal has several subsections, want Agricultural Suitability and Potential Yields. Not measures that depend on human activity, as those will be endogenous to development. Flash application to select data series and download georeferenced data.

- Data for 49 crops, times four input levels, times five water supply types = 980 datasets. Do only for current climate conditions.

- Limited number of crops. Start with wheat, rice, maize, millet, sorghum, barley, oats, rye

- Want crop suitability index at this point at the lowest level of disaggregation (grid-cell level).

The FAO also has a general spatial data website available:

http://www.fao.org/geonetwork/srv/en/main.home

where other subjects may be found. Not sure what is necessary from here at this point.

Frost area from William Masters, in Data folder.

1.4   Ethnic Group Data

There is an Ethnographic Atlas, originally by Murdock. An excel file and codebook are in the Data folder. Tedious, but necessary to import data into Stata, and build up variable definitions.

A subset of the cultures in the Ethnographic Atlas have been expanded upon in detail. These are the SCCS cultures (standard cross-cultural sample). Whoever manages this data is a disaster, and the web sites appear to be the product of someone dumping information without thought. It does appear to be available in R and SPSS (in the Data folder), along with a codebook. Sites for accessing this information are:

http://eclectic.ss.uci.edu/~drwhite/worldcul/SCCSarticles.htm

http://eclectic.ss.uci.edu/~drwhite/courses/index.html

http://intersci.ss.uci.edu/wiki/index.php/SCCS

http://capone.mtsu.edu/eaeff/downloads/SCCScodebook.html

Alesina's replication data gives an example of using SCCS and Ethnographic data to do a similar analysis.

1.5   Concordance

Major project is linking up Ethnographic Atlas information with ethnic boundary data. Use language mapping, as in Alesina? Not a major concern at this point.

1.6   What Do We Want?

Datasets that link economic outcomes, ethnic group characteristics, and agricultural data

- Grid-cell level: lights and urban/rural population data, membership in which ethnic group, membership in which country, agricultural data at the grid level

- Ethnic-group level: summary data on lights and urban/rural data (mean, variance), membership in which country, agricultural data summaries (mean, variances) for that area

- Country-level: summary data on lights and urban/rural data (mean, variance), agricultural data summaries (mean, variances)

Want the ability to study the role of specific crop suitabilities (wheat vs. rice) at the grid-cell level. Is there variation in those cells in lights or urbanization or population size based on crop type, even while controlling for overall agricultural productivity?

Then, can look at ethnic group characteristics, and ask whether some of their particular aspects (taxation, cultural practices) are associated with crop types they use.

So want code that will (clearly and with good documentation) roll up the various sources of geographic grid-cell data into those datasets. In some sense, my goal is to have a clear system set up so that adding a new data source will be relatively easy - copy the existing methods but use new inputs.

Prefer to do this in automated fashion, so using Python to do GIS work directly, rather than involving pointing/clicking in GIS application.

The three datasets listed above are the long-run goal. At this point I want to get a minimum working example of how to do some of this. So think of this MWE:

- Include crop suitability data for rain-fed wheat using low inputs from GAEZ

- Include night light data for 2012

- Include urban population from HYDE

- Roll up to country-level totals, means, and variances for each of the three series.