

# Preparing Data and Running Simple Regressions

**Due April 1st**

For this assignment, you will need to do the following:

1. Edit the “Class-Reference.r” script to perform the rasterization of the DMSP lights data
2. Edit the “Class-Reference.r” script to get the centroids of GADM districts and write those centroids to a dataset
3. Edit the “Class-Lights.r” script to resample the DMSP data to match the GADM extents perfectly
4. Run the “Class-Reference.r” and “Class-Lights.r” scripts to create the baseline datasets of mean lights per district
5. Create a new script that does zonal statistics for GADM districts on the data for number of frost-free days. This data is in the GAEZ folder, a file called “res01-lt3-crav6190.tif”. You will need to make sure extents and resolutions match with the GADM data.

Once you have done that, you will need to combine this data into a single dataset and perform some simple regressions. I do not care if this is in R or in Stata. The script you use should do the following:

1. Merge the data on night lights and frost-free days, by district
2. Create a new variable on log night lights, adjusted for 0 values as in the Henderson et al paper discussed in class
3. Create a new variable of 0/1 for the presence or absence of lights
4. Run regressions of log night lights (by district) on frost days (by district) without any fixed effects, with country fixed effects, and with province/state fixed effects. Produce a table showing the estimated effect of frost days, and the R-squared.
5. Run regressions of 0/1 lights (by district) on frost days (by district) without any fixed effects, with country fixed effects, and with province/state fixed effects. Produce a table showing the estimated effect of frost days, and the R-squared.