**Steps 2024-5-12**

@Zeyu: I have shared a DB folder “ra\_biodiversity” with you. Please find data and put your work there.

**1. General idea**

The goal of this project is to quantify the impact of large industrial plant opening events on ecosystem outcomes. We will follow the general research framework use in Kim (2020) (see “./kim\_2020.pdf”). Basically, we will use the “Million Dollar Plant” list as a quansi-experiment, comparing ecosystem outcomes in counties where new plants opened (“winners”) and the plants’ runner-up choices (“runner-ups”), before vs after plant opening. A standard DID design.

A map of the united states with red circles

Description automatically generatedA graph with a line and a line

Description automatically generated

The two figure above in Kim (2020) provides good intuition. For the map on the left, blue circles are places with actual plant openings, and red dashed circles are runner-ups. These are the treated and control groups. On the right are DID graph, where time “0” is the year when the treated plants actually opened. Blue line represents treated counties, dashed line is control.

**2. Data task: Process winners and runner-ups location information**

The plant opening winner and runner-ups county data are included in “./data/plant\_opening/MDP\_1980\_2008.xlsx”. For each observation, your task is to find its “FIPS county code” (a 5-digit code that uniquely identifies counties in the US), based on state and county name information.

Specifically, this is what the xlsx file looks like:

A screenshot of a computer

Description automatically generated

Using State and County column, match with FIPS code using Census 2010 county shapefile data contained in “./data/shp\_county\_2010”. This is what the shapefile data looks like.

A screenshot of a computer

Description automatically generated

For example, for the first county, the name of the county is “Cleburne”, and its 5-digit FIPS code is “01029”, which is 2-digit state code followed by 3-digit county code. A tricky thing here is that the shape file does not have State abbreviation. For a state-code to state abbreviation crosswalk, see here: <https://www.bls.gov/respondents/mwr/electronic-data-interchange/appendix-d-usps-state-abbreviations-and-fips-codes.htm>

Once you are done with the FIPS code matching, we can merge in ecosystem outcomes and run DID.