**Steps 2024-6-10**

@Zeyu: For next steps, let’s combine the Million Dollar Plant and bird diversity data to conduct some preliminary DID analysis. In particular, we are interested in whether plant opening reduces biodiversity outcomes (such as bird abundance and richness), relative to runner-up counties.

**1. BBS bird biodiversity DID graph**

The following graphs are, again, from Kim (2020) (see “./kim\_2020.pdf”). Let’s try repeat the right-hand side chart, but with county-year level bird outcomes that you derived from the BBS.

A map of the united states with red circles

Description automatically generatedA graph with a line and a line

Description automatically generated

Let’s consider two outcome variables:

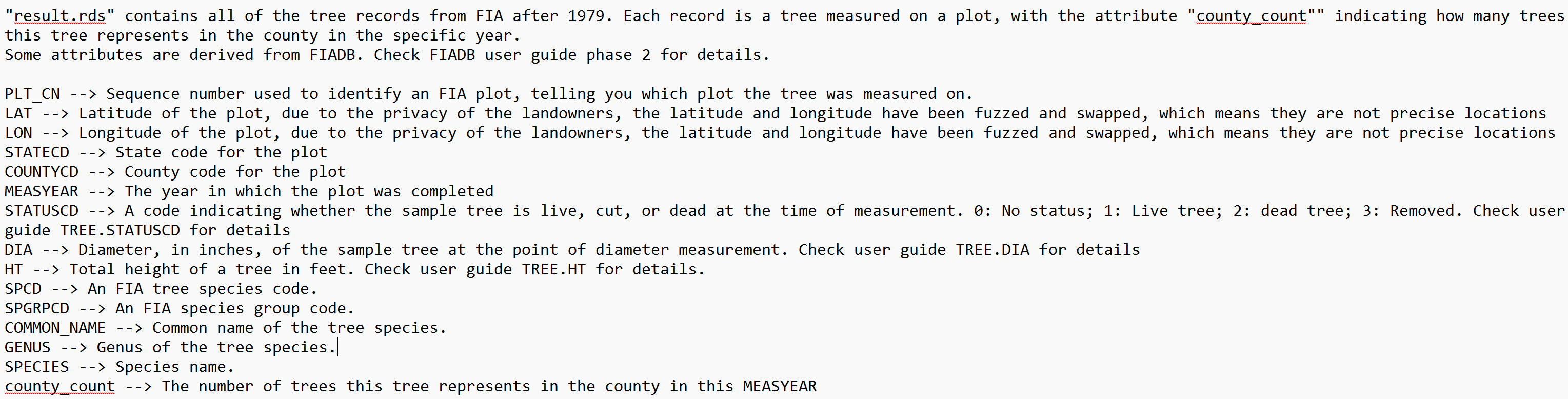
- log of bird abundance in a county-year

- log of bird richness in a county-year

The DID analysis is simply plotting average outcome in winner counties (solid) vs runner-up counties (dash) from 4 years before to 5 years after “year 0” which is the plant opening year.

**2. Data task: Process Xiuyu’s FIA tree data to build county-year level tree abundance and richness measures**

The data file Xiuyu sent seems to be at the individual tree level. Let’s aggregate that data to the county-year level, and compute abundance and richness for trees, just as we did for birds. I repeat Xiuyu’s data description below:



Ultimately we want two biodiversity measures:

- Abundance: total number of tress in the county-year. This should come from the “county\_count” field

- Richness: total number of different kinds of trees in the county-year. This should come from the “SPCD” field

The final dataset should look something like:

|  |  |  |  |
| --- | --- | --- | --- |
| County FIPs | Year | Abundance | Richness |
| 06032 | 1995 | 50 | 3 |
| 06032 | 1996 | 55 | 4 |
| … | … | … | ... |

Once we have this data, we can merge it to the Million Dollar Plant dataset and conduct DID just as we did with birds outcomes.

**Outcomes**

**1. BBS bird biodiversity DID graph**

Bird log(abundance) did:

**A graph with lines and numbers

Description automatically generated**

Bird log(richness) did:

**A graph with a line and a line

Description automatically generated**

**Concerns and notes:**

1. Since we cannot guarantee that county\_x\_year\_panel\_data contains all the FIPS codes that mdp\_data does, we have to give up tracking some of the plants. As a result, we have 310 distinct plants in total, but only 197 plants after joining the two datasets. However, I think this is a reasonable loss since the US has 3,242 FIPS codes in total, and we only have 2,013 FIPS codes in our county\_x\_year\_panel\_data. Therefore, I think it is reasonable to lose 1/3 of the plants.

2. We do have 7 NA counties in mdp\_data:

1. European Aeronautic Defense and Space Co.(EADS): Berkeley(45015)/ Charleston (45019)/ Dorchester (45035) -- I think EADS did not choose SC for their plants. According to Airbus’s website, they do not have any facilities located in SC. I will choose 45015 as the FIPS, since it covers most of the North Charleston area.
2. Google: Fairfield (45039)/ richland (45079) -- We can’t find Google in SC. However, I found that most of Blythewood city is located within Richland County.
3. Bristol-Myers Squibb: Middlesex (25017)/ Worcester (25027) – According to Google Maps, it has two locations, and one is in Boston. However, the data indicates it is located in Devens, so I use Worcester as the FIPS.
4. MCI: Cook (17031)/ DuPage (17043) – Hard to trace the information, but I do find a company named MCI located in DuPage
5. Racal-Milgo: Hialeah– Dade ---- 12086
6. MCI Communications: Miami – Dade ---- 12086
7. Level 3 Communications: San Mateo (06081)/ San Francisco (06075)/ Marin (06041) – I chose SF as the FIPS since SF is located in the middle of them.

3. The biggest concern would be the volume of the dataset. In the beginning, we have 330 plants in total. After joining the data, the number of plants dropped from 330 to 197 (verified number). Additionally, the years from `county\_x\_year\_panel\_data` do not contain all the data from 1966 to 2021, as some years may be skipped. Then, I filtered the data again to ensure every `plant\_id` corresponds to 10 data points. At this step, we only have 85 plants left.

A close-up of a sign

Description automatically generated

**2. FIA tree biodiversity DID graph**

Tree log(abundance) did:

A graph with lines and numbers

Description automatically generated

Tree log(richness) did:

A graph with red and blue lines

Description automatically generated

**Concerns and notes:**

1. Fortunately, we have 3,114 distinct FIPS codes in the tree data and only lost 10 FIPS after joining the two datasets. Therefore, the number of plants dropped from 310 to 292.

2. Unfortunately, the years in the tree data only cover from 1980 to 2024. Therefore, after filtering the data to keep only plants with 10 years of data, the number of plants dropped from 292 to 15.

**Other thoughts about the method:**

The method of combining datasets might be problematic, whether for birds or trees. I merged the datasets by years and FIPS codes. However, this poses a problem if the winning and runner-up plants are located in the same county and were publicized during a similar period. Since we have 310 distinct plants but only 224 distinct FIPS codes, the DID might not work effectively in situations where there is biodiversity in both the winning and runner-up groups. I believe this is one of the reasons the graphs are not as good as we expected.

Second, the number of plants with sufficient information (spanning 10 years) might be too small, especially for the tree analysis.