PS6

Becky Crouse

March 21, 2023

Data Cleaning

I'm utilizing ISORA data for this project. ISORA, or International Survey on Revenue Administration, is an international survey of revenue agencies which gathers metrics on taxpayer audits, assessments, and agency staffing. I utilize three data sets (5, 7, and 10) found online here: `https://data.rafit.org/?sk=8b008788-ebde-4d61-bc90-7438d6aa12dc&sId=1637191076670`.

    Because each of these files is formatted slightly differently, I read in and clean each file separately by performing the following steps:

1. Read in the data and subset it to remove upper rows and extra columns that I'm not using.

2. Adjust the column headings to reflect the first row headings.

3. Repeat above for each data set. Because two of the files have two sets of information that I want, I end up with 5 data frames.

4. I then convert each of the "wide" data frames to "long" data.

5. I use SQLite in R to merge each of the "long" data frames together into a master data frame.

6. I create new variables using the combined data.

7. Finally, I do my final review and cleaning of the data. I remove any observations that are not readily interpreted:

   - Rows with missing information
   - Countries reporting 0 taxpayers
   - Countries reporting more audits than active taxpayers (audit rate >1)

Data Visualizations

My interest in this data is in understanding how tax authorities differ on their audit aggressiveness as well as the taxpayers response. I created three visualizations to help me understand these relationships.

    First, I created scatter plots to visualize the data. I plot the audit rate (calculated as number of audits over the number of active taxpayers) against the total government revenue as a percent of GDP.

Government Revenues and Tax Audit Rates
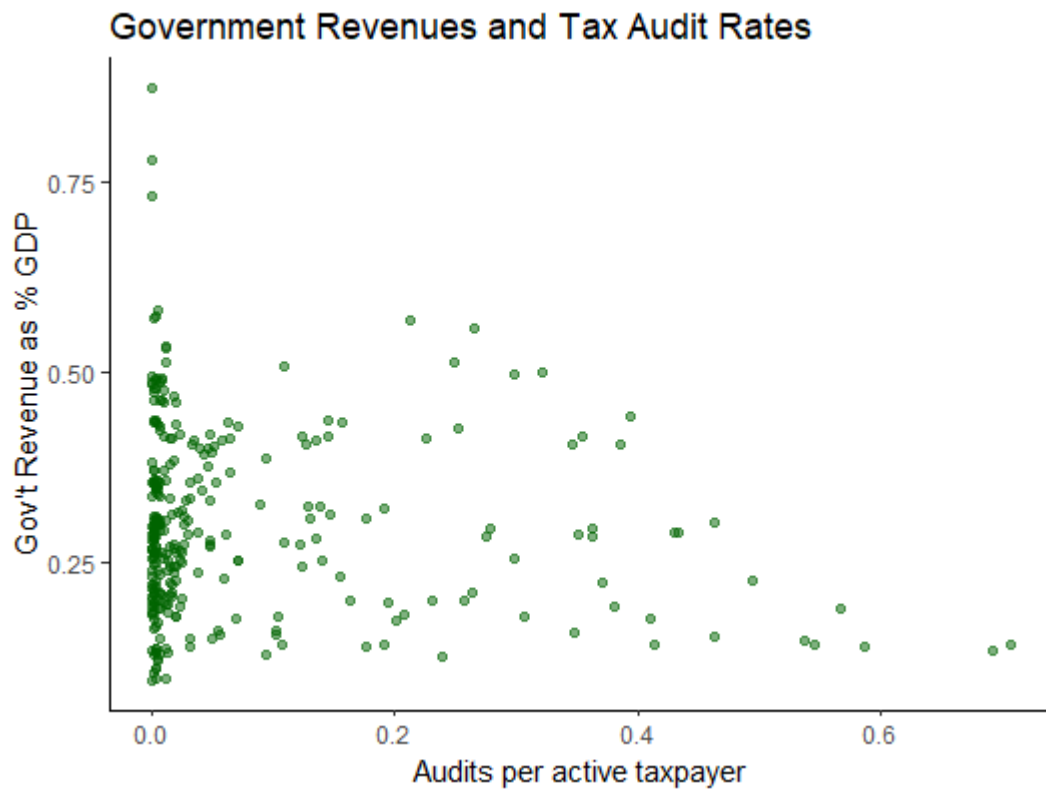
I then created a similar scatter plot, this time showing the rate of audits with assessment (calculated as number of audits with assessment over the number of active taxpayers) against the total government revenue as a percent of GDP.

Government Revenues and Tax Audit Assessment Rates
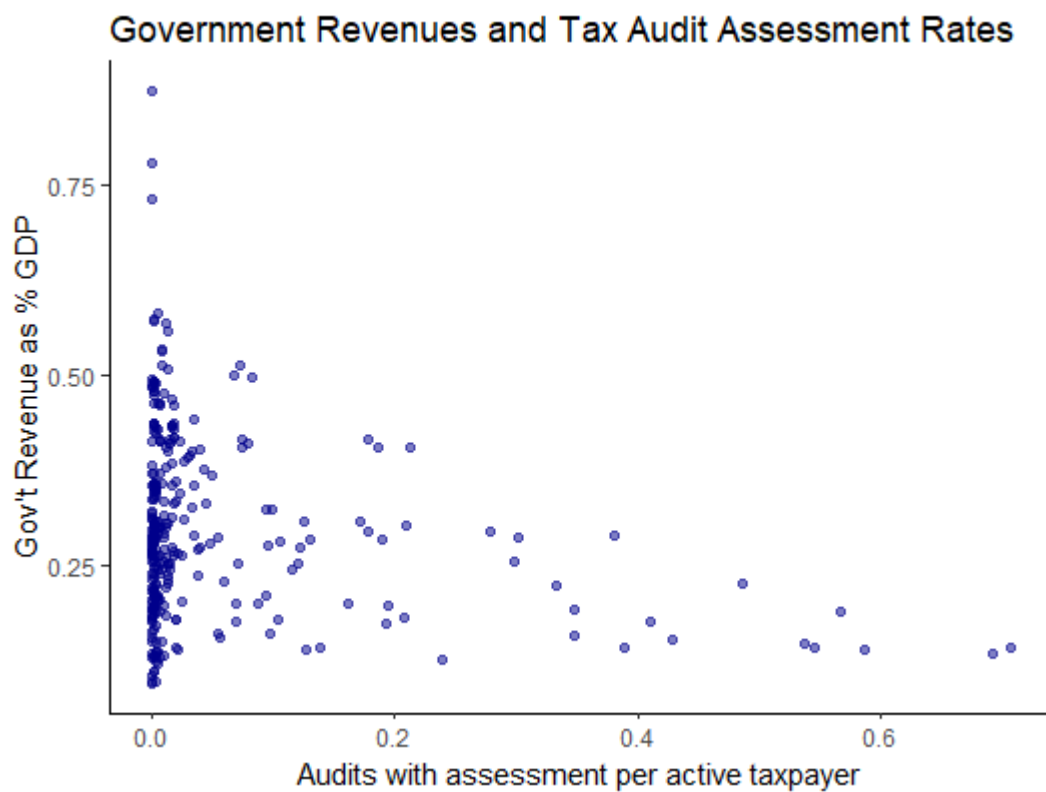
Finally, to understand the relationship between audit assessments and government revenue, I plotted a Loess curve of the regression of government revenue as a percent of GDP on audits with assessment. As expected, I find a significant

negative relationship between these two, suggesting that audit assessments increase as government revenue falls (or vice versa). In untabulated analysis, I find a similar, though less significant, relationship between the overall audit rate and government revenues.

Government Revenues and Tax Audit Assessment Rates