

This week we will focus on Davidson County: looking at the spatial pattern of crime incidents; examining the distribution of a few types of businesses and government services; and taking a first look at local real estate markets.

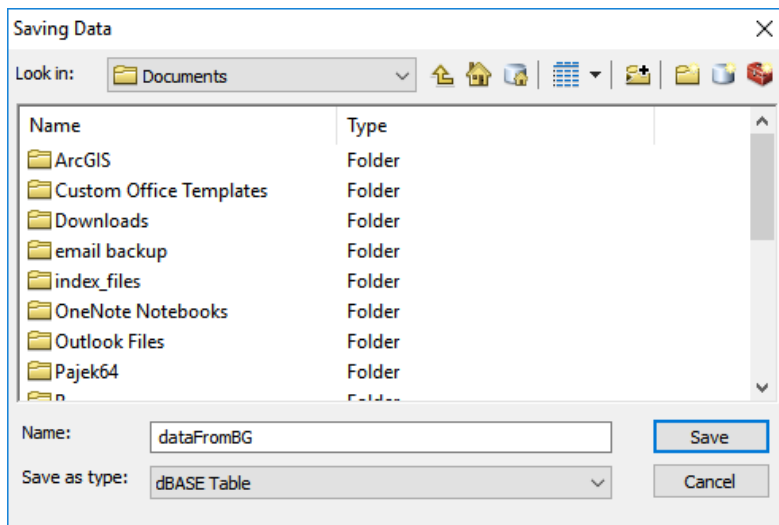
*Open Data Portals* are one of the most valuable sources of local spatial data for large American cities. I made the following shapefiles from csv files found at [Nashville's Open Data Portal \(https://data.nashville.gov/\)](https://data.nashville.gov/):

- *S:\TEFF\450\2020\arcview5\MNPDinc.shp* (locations of MNPD incidents 2019)
- *S:\TEFF\450\2020\arcview5\BldgPermits.shp* (locations of projects with building permits)
- *S:\TEFF\450\2020\arcview5\PrprtyViol.shp* (locations of properties cited for upkeep violations)
- *S:\TEFF\450\2020\arcview5\ShrTrmRntl.shp* (locations of licensed short-term rentals)
- *S:\TEFF\450\2020\arcview5\BeerPermits.shp* (locations with Beer license)
- *S:\TEFF\450\2020\arcview5\CommunityResource.shp* (locations of resources for the disadvantaged)

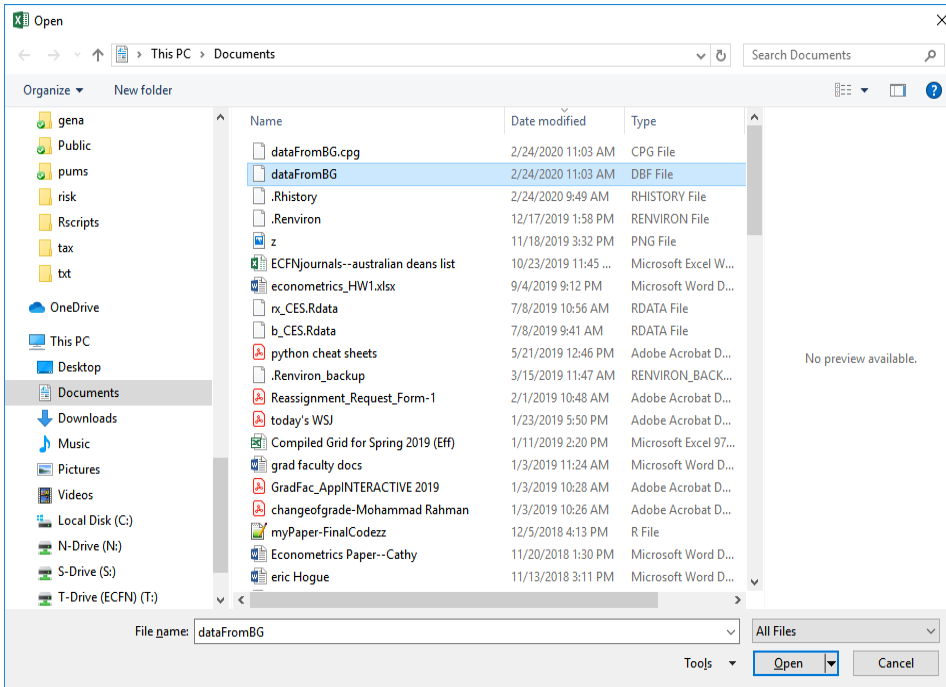
*ReferenceUSA* is good source containing data for US business locations. These are proprietary data, but are available to MTSU students through the library database links. To access these data, go to <http://www.referenceusa.com/Home/Home> and then click "U.S. Businesses" and then the "Advanced Search" tab on the search box. In the resulting search page, one can narrow down the geography to a specific county or zipcode, and narrow down the type of business to a particular SIC or NAICS code. One can typically download only 250 records per search, so narrow searches are necessary. I made the following shapefiles from csv files downloaded from ReferenceUSA:

- *S:\TEFF\450\2020\arcview5\Davidson\_bars.shp* (locations of bars, nightclubs, pubs, and cocktail lounges)
- *S:\TEFF\450\2020\arcview5\liquorStores.shp* (locations of liquor stores)
- *S:\TEFF\450\2020\arcview5\highRloan.shp* (locations of pawn, title loan, or payday loan establishments)

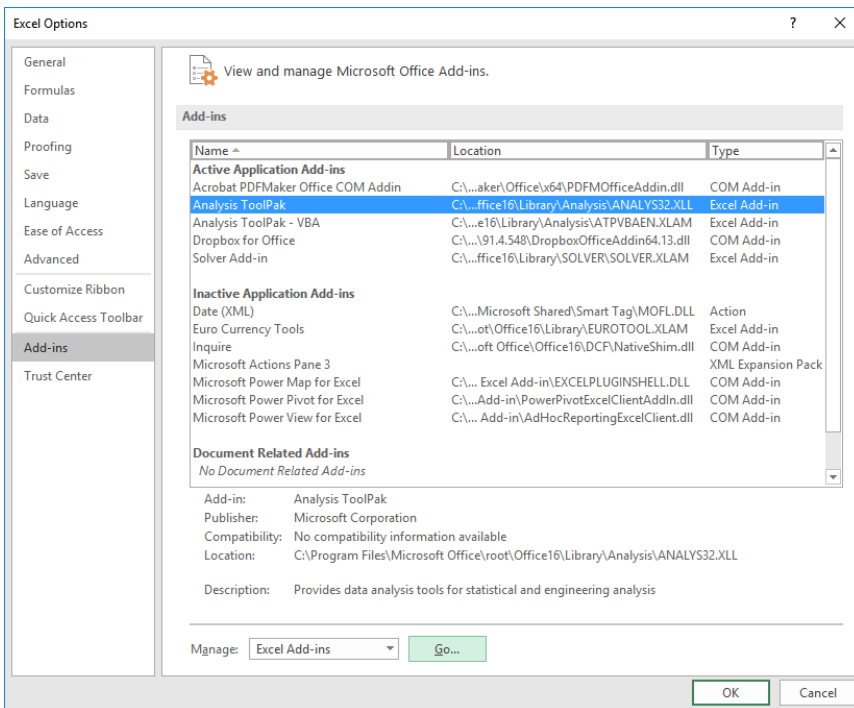
**In-class work:** Load the shapefile *S:\TEFF\450\2020\arcview5\BG\_Davidson.shp* and 'sheet 1\$' in the Excel file *S:\TEFF\450\2020\arcview5\BGcrimeData.xls*. Join the Excel data to the attribute table of the shapefile using *geoid* as the join field for both tables.



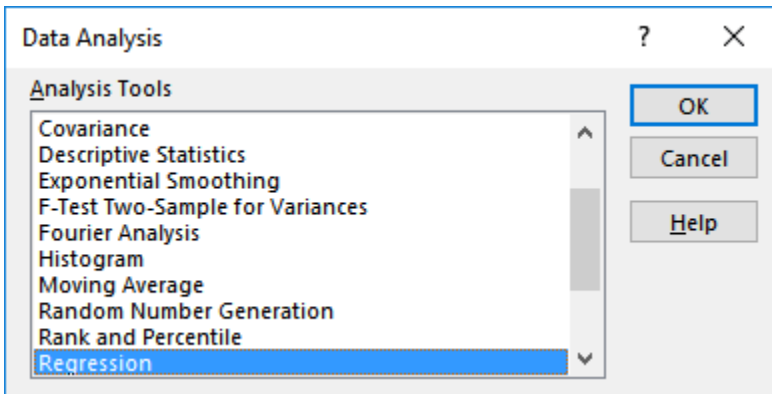
Once successfully joined, export the attribute table of the shapefile (right click on *BG\_Davidson* in the *ToC*, open the attribute table, click on the icon at the top far left of the attribute table, click on *Export* on the drop-down menu, click on *Export* on the drop-down menu, click on the browse button to the right, select 'dBASE Table' in the *Save as type* box, select your *Documents* folder as the location, and give the exported file a name of your choosing). Save, but do not bring into the view. Next, open Excel, and load your exported file. The file has extension \*.dbf, which Excel considers a foreign filetype, so you must select 'All Files' in the drop- down menu to the right in order to find it.



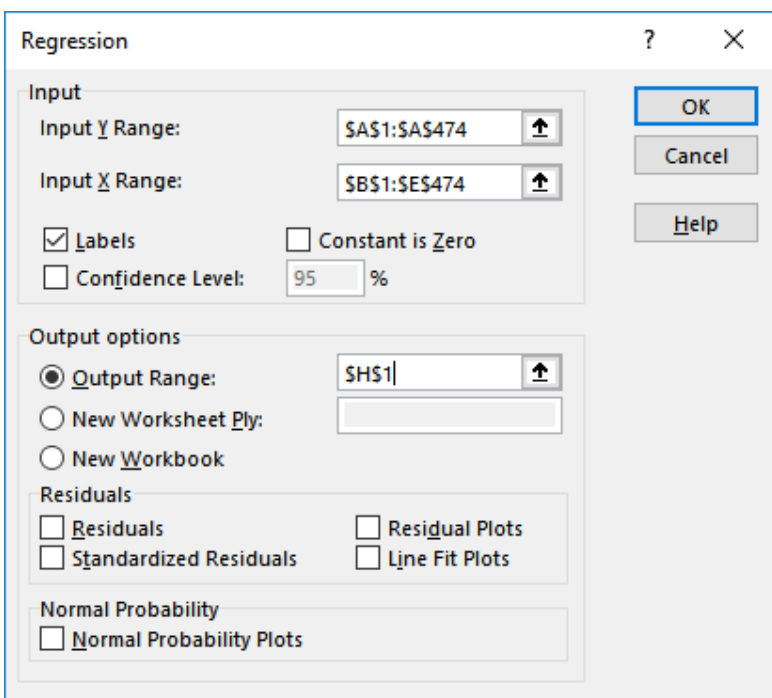
Once loaded, save the file in Excel Workbook format. We will attempt a linear regression model to see if the number of crime incidents influences the presence of short-term rentals. We must first load the Excel Analysis ToolPak containing functionality for linear regression. Click on the 'File' tab in the Excel ribbon, and then click on 'Options' in the sidebar menu. In the Excel Options pop-up box, click on 'Add-ins' in the sidebar menu, highlight 'Analysis ToolPak' in the Add-ins box, and click the 'Go...' button at the bottom.







In the 'Regression' pop-up box, select the dependent variable column for the 'Input Y Range' and all of the independent variable columns as the 'Input X Range'. Be sure to include the header row when selecting your ranges, and put a check mark in the 'Labels' box. For 'Output Range' simply select a cell at the top to the right of your data. Click OK.



### Interpreting the output:

1. **R square** is the proportion of the variation in the dependent variable explained by the model. This will lie between 0 and 1.
2. The **Coefficients** are important for their signs:
  - a. A negative value indicates that the dependent variable and that particular independent variable move in opposite directions (e.g., if the independent variable decreases, then the dependent variable will increase).
  - b. A positive value indicates that the dependent variable and that particular independent variable move in the same direction (e.g., if the independent variable decreases, then the dependent variable will decrease).
3. The **P-value** gives the probability that the true value of the coefficient equals zero (implying that there is no relationship between the dependent variable and the independent variable). By convention, we consider any probability greater than 0.10 to mean that there is no relationship between the dependent and independent variable.

**Homework:** There are two parts to this homework.

Make heatmaps for the distribution of different types of crimes, as well as the distribution of the business-related shapefiles. Pick point layers that you think might have an influence on the distribution of short-term rentals. Do around 20 heatmaps, and pick out five that most look like they have a relationship with short-term rentals. Export the maps and put these in a PowerPoint presentation.

Build a model explaining the spatial distribution of short-term rentals. Use as independent variables those variables you identified in your five maps above, and a few others you think might be interesting. Put your regression results in your PowerPoint and be prepared to explain what they mean.