Heather Trantham

Capstone 1 project

The data chosen for my Capstone 1 project is data is from the US Census Bureau. The data contains information on commuter travel in the United States. The file was compiled from STF-S-5, Census of Population 1990: Number of Workers by County of Residence by County of Work. For the six New England States (CT, ME, MA, NH, RI, VT), data are provided for Minor Civil Divisions (MCDs) instead of for counties.

The data was a table of fixed-width formatted lines. This was read into a data frame using the pd.read\_fwf. It was necessary to specify the width of each column before the file was read in. The column labels were subsequently added to the data frame.

Each row in the data frame contains information about the state and county where the worker resides, the state and county where the worker works, and the number of workers starting in the county of residence and working in the destination county. The states and counties are identified in the data frame by the Federal Information Processing Standard (FIPS) code for each state and county.

In order to make future plots more readable, the two letter state abbreviations were defined in a dictionary that was then mapped to the FIPS\_state\_res column of the dataframe.