create-analysis-dataset.R

1. Load 2019 GA Zip Code shapefile (from Howard)
   1. 697 Zip Codes
   2. Filter out 12 Zip Codes with 0 population
      1. 30118, 30111, 30320, 30334, 30069, 30149, 30751, 30597, 31698, 31304, 31760, 30609
   3. Result: 685 2019 GA Zip Codes with population data and polygons
2. Load 2019 GA county shapefiles (from Tigris)
   1. 159 Counties
   2. Filter to 5 ATL counties (Fulton, DeKalb, Gwinnett, Cobb, and Clayton)
   3. Convert projection to WGS84 to match ZIP code file
3. Identify GA Zip Codes which intersect the 5 ATL county area
   1. Result: 130 zip codesA map of a city

      Description automatically generated
4. Load Daily ED visit counts by zip code for 2005-2020 (from Howard)
   1. Filter to only include records for 130 zip codes in ATL area
      1. 54,823,247 🡪 760,268 records
         1. Results come from 130 zip codes (5,844-5,857 days for each one)
            1. Double entries for 548 zip-years
5. Load Air pollution data (from Howard)
   1. Years provided: 2005-2019
   2. Filter to only include records for 130 zip codes in ATL area
      1. 3,805,751 🡪 708,489
         1. Results come from 130 zip codes (all but 3 have 5,478 days)
            1. 30009: 4383 days, 30046: 3652 days, 30363: 4748 days
6. Load meteorology data (from Howard)
   1. Years provided: 2005-2018
   2. Filter to only include records for 130 zip codes in ATL area
      1. Total of 661,039 records
         1. Results come from 130 zip codes (all but 3 have 5113 days)
            1. 30009: 4018 days, 30046: 3287 days, 30363: 4383 days
7. Load time data (from Howard)
   1. Years provided: 1993-2020
      1. 1 record per day from this period = 10227 records