How Close is Close? The Spatial Reach of Agglomeration Economies

Online Appendix: Data Sources

Stuart S. Rosenthal and William C. Strange

This file provides additional detail on the data sources associated with Figures 1-4 in the paper. Portions of this information also appears in the ReadMe file posted to the data archive along with programs and that portion of the data that can be shared.

The primary data used for the project are Dun & Bradstreet establishment level files from the D&B Million Dollar Database. These data were obtained through the Syracuse University data library which has a site license for the data. The terms of the license do not allow us to post or share the data. However, many other universities and public libraries have similar site licenses.

Additional data for Figures 1-4 include shape files for the maps in those figures. This includes a state boundary map, a county boundary map, a map of primary roads, a map of local streets, and a map of the buildings in the five boroughs of New York City. These boundary files were obtained from publically available sources.

Details on each of the data sources are below.

Dun & Bradstreet establishment data

The D&B data were obtained through the Syracuse University data library with the Syracuse site license. Access to the data will require similar access to a site license at other university and public libraries.

The data are from the Dun & Bradstreet Million Dollar Database, details of which can be found at the URL below:

https://www.mergentmddi.com/index.php/search/iplanding

The data provide establishment level information on nearly all private and public companies in the United States and their executives. This includes various measures of a company's profile and activity in addition to the street address of the establishment and the latitude and longitude of its location. Measures include employment, SIC and NAICS codes, sales, corporate status of the establishment (headquarters, branch or single site), and much more.

These data were accessed through the Syracuse University data library at the following URL:

https://www.mergentintellect.com/index.php/search/index

GIS boundary files for Figures 1-4

Figures 1-4 use a total of five different boundary files. Figure 1 uses a state boundary map in addition to a map of primary roads. Figure 2 also uses the state boundary map. Figure 3 uses a county boundary map and a map of local streets in New York City. Figure 4 uses a map of buildings in Manhattan.

The boundary files for Figures 1-4 are all available through public sources. This includes:

- State and county boundary files

  State and county boundary files can be downloaded from the US census department website for "Cartographic Boundary Files – Shapefile" at:

  https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html

  This site provides boundary files for many different geographic areas in the United States. Files used for this paper were obtained by downloading "cb_2018_us_state_5m.zip" for the state file and "cb_2018_us_county_5m.zip" for the county file.

- Primary road boundary file

  The boundary file for primary roads in the United States can be downloaded from the US Census Bureau "TIGER/Line Shaprefile, 2016, nation, U.S., Primary Roads National Shaprefile" at:

  https://catalog.data.gov/dataset/tiger-line-shapefile-2016-nation-u-s-primary-roads-national-shapefile

  This site provides several files. The file used for this paper was obtained by downloading the file "Shapefile Zip File".

- Local streets boundary file for New York City

  The boundary file for local streets in New York City can be downloaded using the "Zip file" from the following URL:

  https://catalog.data.gov/dataset/nyc-street-centerline-cscl

  This file is made available and posted by the Department of Information, Technology and Telecommunications in the Department of City Planning in New York City.

- Buildings in the NYC

  The boundary file for specific buildings in New York City was obtained from the MapPLUTO website posted by the New York City planning department. The shape file can be downloaded from the following URL:

https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page#mappluto

From this page, the shape file was obtained by downloading the file "MapPLUTO - Shoreline Clipped (Shapefile)".

Data References

Dun & Bradstreet. 2019. "The Million Dollar Database." Mergent Intellect by FTSE Russell. https://www.mergentmddi.com/index.php/search/iplanding and https://www.mergentintellect.com/index.php/search/index (accessed May and June, 2019).

United State Census Bureau. 2018. "Cartographic Boundary Files – Shapefile." United States Census Bureau. File "cb_2018_us_state_5m.zip" at https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html (accessed May, 2019).

United State Census Bureau. 2018. "Cartographic Boundary Files – Shapefile." United States Census Bureau. File "cb_2018_us_county_5m.zip" at https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html (accessed May, 2019).

Department of Information, Technology and Telecommunications, New York City Department of City Planning. Updated July 1, 2020. "NYC Street Centerline (CSCL)." File "Zip file" at https://catalog.data.gov/dataset/nyc-street-centerline-cscl (accessed May, 2019).

New York City Department of City Planning. Updated June, 2020. "MapPLUTO - Shoreline Clipped (Shapefile)." https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page#mappluto (accessed May, October, 2019).