County Buddy - Contextualizing Outlier Data in Institutional and Jurisdictional Enclaves

Colin Vu[a], Second Author Name[b]

[a]College of Computing, Georgia Institute of Technology, Atlanta, Georgia, USA

[b]Department, University, City, Country

1 Purpose

County Buddy is a sophisticated tool designed to allow users to extract specific, curated insights from outlier data at the county and census tract levels. Its primary objective is to provide contextual understanding of whether observed outliers can be attributed to the presence of institutions or jurisdictional enclaves within these regions. The tool compiles data across several key demographic categories, including **incarcerated population**, **student population**, **military population**, and **Native American population**. County Buddy identifies counties or tracts exhibiting outlier data and informs users of the potential influence of nearby jurisdictional enclaves on these anomalies.
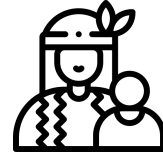

Incarcerated Population   Student Population   Military Population   Native American Population

For example, consider a study focused on regions with a high prevalence of a newly emerging viral disease. The researcher may identify a census tract exhibiting an unusually high infection rate, in contrast to the surrounding tracts, which show typical rates. In this case, County Buddy could inform the user that the outlier data point may be caused by a prison within the county, where the disease is prevalent but confined to the institution, thus remaining isolated from the surrounding area.

Category	Total Number of Institutions	Total Population of Institution	Counties Affected
Prisons	6737	1978489	2656
Universities	3541	2794201	1218
Military Institutions	743	328615	514
Native American Reservations	611	3745005	415

Table 1: County Summary Statistics

Category	Total Number of Institutions	Total Population of Institution	Tracts Affected	Nearby Tracts Affected
Prisons	6737	1978489	4991	
Universities	3541	2794201	2577	7819
Military Institutions	743	328615	2074	
Native American Reservations	611	3745005	1445	

Table 2: Tract Summary Statistics

2 Usage

County Buddy is provided in the form of two CSV-format spreadsheets: one containing county-level data and the other containing census tract-level data. Each record corresponds to a distinct county or census tract. Within each record, any demographic data categories (such as university student population, military population, etc.) are excluded or "zeroed out" if they do not meet the criteria for being considered outliers. Outlier data is defined as data where the percentage of the total population represented by a given demographic exceeds three standard deviations above the mean.

FIPS	State	County	Total Pop.	Prison Pop.	% Prison	Institutions	...
12043	Florida	Glades County	12126	1640	13.50%	FLORIDA ENVIRONMENTAL INSTITUTE; MOORE HAVEN CORRECTIONAL FACILITY; GLADES COUNTY DETENTION CENTER	...
12045	Florida	Gulf County	14192	0	0.00%		...
12047	Florida	Hamilton County	14004	2323	16.50%	HAMILTON WORK CAMP; PANTHER SUCCESS CENTER; HAMILTON CORRECTIONAL INSTITUTION; HAMILTON CORRECTIONAL INSTITUTION ANNEX; HAMILTON COUNTY JAIL	...

Table 3: Example County-Level Incarceration Data

Below is an example implementation of County Buddy in the Resiliency Map, created by Arpit Narechania. The map's county-level data has been joined with County Buddy's county-level data using their FIPS codes, and County Buddy's data is only shown when not "zeroed out."
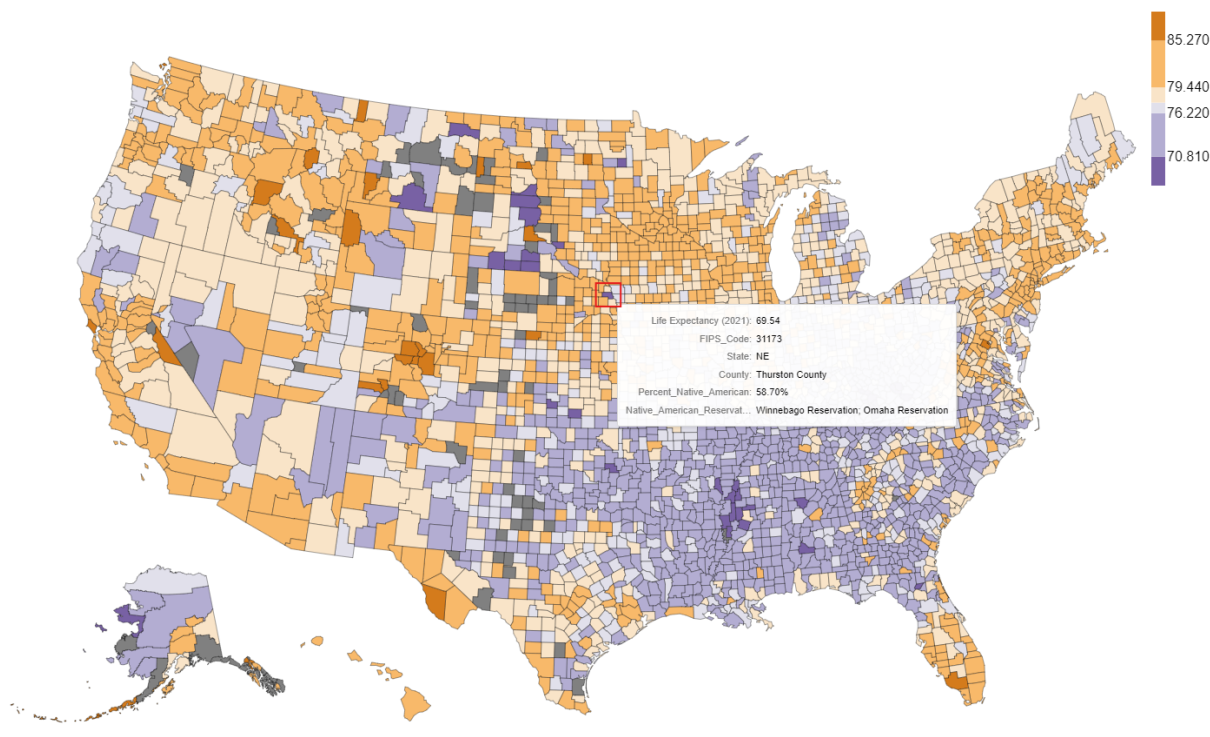

Figure 1: Example usage of County Buddy in https://ocular.cc.gatech.edu/resiliency-app. The figure shows Thurston County, NE, having an abnormally low life expectancy. County Buddy informs the user that this may be the result of the presence of the Winnebago and Omaha Native American Reservations, as the Native American population is about 59% of the total population. The life expectancy in Thurston County is significantly lower than any of its surrounding counties, all of which having Native American populations that do not meet County Buddy's threshold.

3 Data Dictionary

Field	Description / Notes
FIPS code	The county/tract's geographic identifier. For county data, this is 5 digits. For tract data, this is 11 digits.
State Name	Name of the state where the county/tract is located.
County Name	Name of the county where the county/tract is located.
Tract Name (only for tract data)	Name of the tract.
Total Population	2020 population of the county/tract.
Prison Population	Residential population of the county/tract registered as incarcerated by the Census Bureau's 2020 Group Quarters survey.
Prison Population as Percent of Total Population	Percent of the total population that is registered as incarcerated, 2020.
Prisons	Prisons whose physical boundaries are contained within the county/tract.
University Student Population	Residential population of the county/tract registered as a university student by the Census Bureau's 2020 Group Quarters survey.
University Student Population as Percent of Total Population	Percent of the total population that is registered as a university student, 2020.
Universities	Universities whose physical boundaries are contained within the county/tract. Only active public and private non-profit colleges (designated as type 1 or 2 by IPEDS) a student population of more than 500 as of 2020 are considered.
Nearby Universities	For tract data only. Universities whose physical boundaries are contained within the tract, or are within 1 kilometer of the border of the tract. Only active public and private non-profit colleges (designated as type 1 or 2 by IPEDS) a student population of more than 500 as of 2020 are considered.
Military Population	Residential population of the county/tract registered as active military by the Census Bureau's 2020 Group Quarters survey.
Military Population as a Percent of Total Population	Percent of the total population that is registered as in military, 2020.
Military Institutions	All Military Institutions whose borders contain some area within the region.
Native American Population	Residential population of the county/tract listed as Native American by the 2020 Census's demographic survey.
Native American Population as a Percent of Total Population	Percent of the total population that is surveyed as Native American, 2020.
Native American Reservations	All Native American Reservations whose borders contain some area within the region.

Table 4: County Buddy Data Dictionary

References

1. U.S. Census Bureau (2020). *Cartographic Boundary Files.* County level and tract level data and physical boundaries from the U.S. Census Bureau. Retrieved from https://www.census.gov/geographies/mapping-files/time-series/geo/cartographic-boundary.html.

2. U.S. Census Bureau (2020). *2020 Census Demographic and Housing Characteristics File (DHC).* County and tract level demographic data from the U.S. Census Bureau. Retrieved from https://data.census.gov/.

3. U.S. Census Bureau (2020). *Group Quarters Populataion By Major Group Quarter Type.* County and tract level group quarters data from the U.S. Census Bureau. Retrieved from https://data.census.gov/.

4. Narechania, A. *Resiliency App.* Sample visualization with County Buddy data integrated. Retrieved from https://github.com/arpitnarechania/resiliency-app.

5. National Center for Education Statistics (2018-2019) *Integrated Postsecondary Education Data System - Homeland Infrastructure Foundation-Level Data.* Data and geographic locations for Colleges and Universities. Retrieved from https://web.archive.org/web/20250310062140/https://public.opendatasoft.com/explore/dataset/us-colleges-and-universities/table/.

6. U.S. Department of Transportation (2023) *Military Bases.* Data and geographic locations for Military Bases. Retrieved from https://catalog.data.gov/dataset/military-bases1.

7. United States Department of the Interior (2021) *Indian Reservations.* Data and geographic locations for Native American Reservations. Retrieved from https://www.arcgis.com/home/item.html?id=8fded139728f48b3b374a5dbf41dd4ec.

8. Oak Ridge National Laboratory (2017) *Prison Boundaries.* Data and geographic locations for Native American Reservations. Retrieved from https://www.arcgis.com/home/item.html?id=2d6109d4127d458eaf0958e4c5296b67.