County Buddy - Contextualizing Outlier Data in Institutional and Jurisdictional Enclaves

Colin Vu[a], Leila Baniassad[a], Clio Andris[a],

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

1 Purpose

County Buddy is a novel dataset of information on the presence, number of, and institution names of special populations at the U.S. county and census tract level. These special populations include **incarcerated population**, **student population**, **military population**, and **Native American population**.

The objective of this dataset is provide context to bolster narratives about socio-economic indicators, provide a potential 'explainer' for why these values appear, or provide a landscape for situating these values.
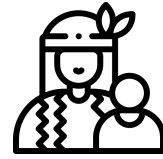

Incarcerated Population   Student Population   Military Population   Native American Population

For example, when studying a newly emerging virus, a researcher may identify a census tract exhibiting an unusually high infection rate and use County Buddy to find that the value may be attributed to a university or a prison within the tract. A map designer may want to help users investigate why they see counties or tracts with much higher average high income or higher birth rate than surrounding areas. By integrating County Buddy into their tooltips or adding radio buttons, the user may learn that this area has a military base or university nearby.

The data can be used by researchers and developers who are studying causality and correlation in health and livelihood outcomes, as they can be easily integrated into statistical models (e.g., in R or STATA). They can also be used by developers looking to enhance their maps of U.S. thematic data (life expectancy, percent smokers, car accidents, number of people on Medicare) with tooltips and extra features.

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 some number of standard deviations above the mean. For county data, two standard deviations are used. For tract data, three standard deviations are used.

Category	Total Number of Institutions	Total Population of Institution	Outlier Threshold (% of Total Pop.)	Counties Affected
Prisons	2,424	1,978,489	8.9%	143
Universities	1,425	2,794,201	4.4%	118
Military Institutions	743	328,615	1.4%	32
Native American Reservations	611	3,745,005	17.8%	60

Table 1: County Summary Statistics

Category	Total Number of Institutions	Total Population of Institution	Outlier Threshold (% of Total Pop.)	Tracts Affected	Nearby Tracts Affected
Prisons	2,424	1,978,489	17.5%	817	
Universities	1,425	2,794,201	19.2%	758	1,030
Military Institutions	743	328,615	8.2%	189	
Native American Reservations	611	3,745,005	15.1%	368	

Table 2: Tract Summary Statistics

FIPS	State	County	Total Pop.	Prison Pop.	% Prison	Institutions	...
12043	Florida	Glades County	12,126	1,640	13.50%	FLORIDA ENVIRONMENTAL INSTITUTE; MOORE HAVEN CORRECTIONAL FACILITY; GLADES COUNTY DETENTION CENTER	...
12045	Florida	Gulf County	14,192	0	0.00%		...
12047	Florida	Hamilton County	14,004	2,323	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."
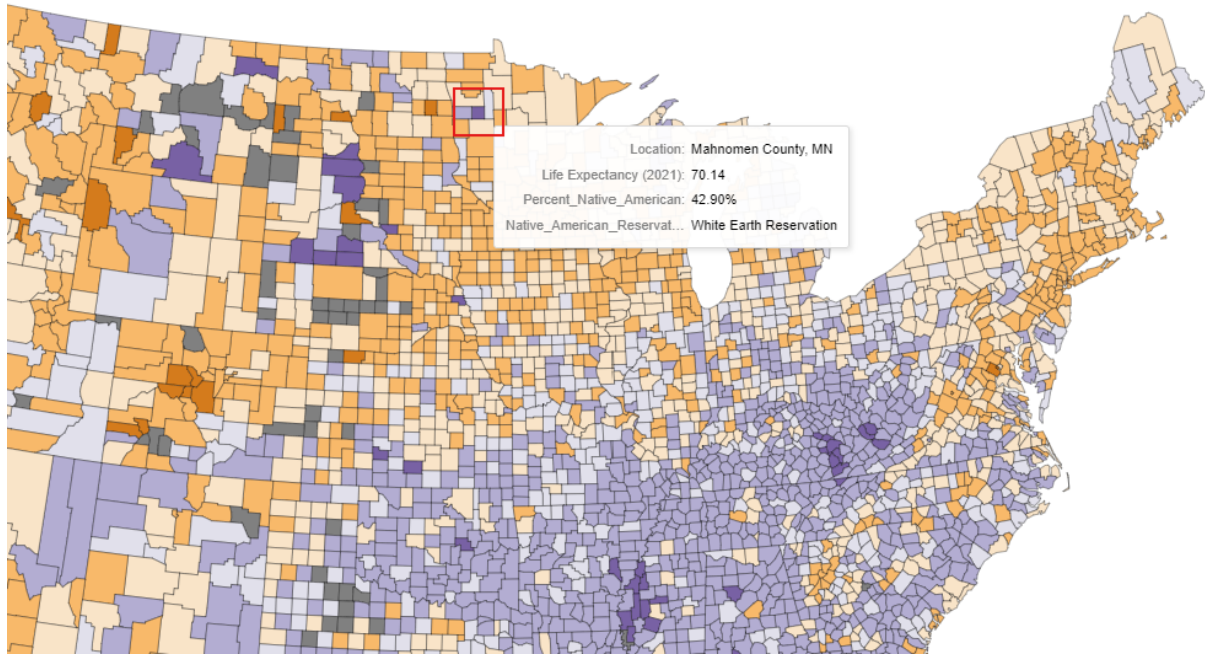
Figure 1: Example usage of County Buddy in https://ocular.cc.gatech.edu/resiliency-app. The figure shows Mahnomen County, MN, having an abnormally low life expectancy. County Buddy informs the user that this may be the result of the presence of the White Earth Reservation, as the Native American population is about 43% of the total population. The life expectancy in Mahnomen County is significantly lower than any of its surrounding counties, all of which have Native American populations that do not meet County Buddy's threshold.

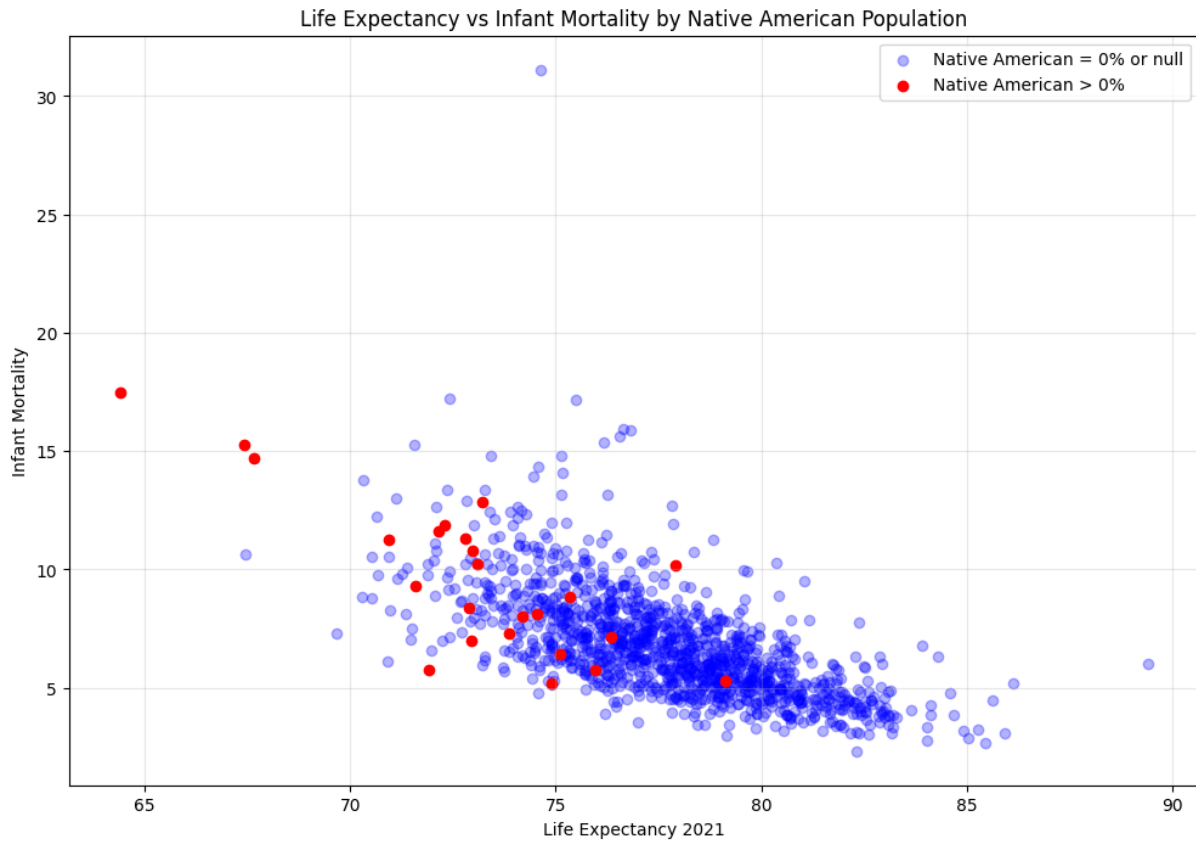Figure 2: Data from https://ocular.cc.gatech.edu/resiliency-app plotted. Life Expectancy rates across counties are compared to Infant Mortality rates. Counties that meet County Buddy's Native American Population threshold are marked in red, whereas the rest are marked in blue.

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. Only prisons with a capacity of 200 or greater in 2017 are included.
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) that are not considered Junior Colleges by NAICS classification and have a student population of more than 700 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) that are not considered Junior Colleges by NAICS classification and have a student population of more than 700 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

4 Disclaimer

Disclaimer about images in this document. The researchers who created this document are not members of the incarcerated, military, or Native American population. If someone from one of these groups would like to help us find a graphic that better represents their population, please reach out.

5 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 secure detention facilities. Retrieved from https://www.arcgis.com/home/item.html?id=2d6109d4127d458eaf0958e4c