



Newsy WIC Investigation Methodological Notes

Newsy based this analysis on data Rachel Gold, Newsy data reporter, obtained through the Freedom of Information Act (FOIA) in October 2022 from the U.S. Department of Agriculture. The data was a list of authorized vendors able to accept benefits from the Special Supplemental Nutrition Program for Women, Infants and Children, better known as WIC. The list included authorized vendors from all 50 states, Washington, D.C., tribes and territories. The vendors were authorized and registered by WIC health departments through the Food Nutrition Service (FNS) system — which the USDA uses to monitor WIC vendors — every year, 2011 through 2020. Newsy reporters limited the analysis to vendors in states (as opposed to territories) and the District of Columbia because U.S. territories do not have reliable ZIP code data. Newsy reporters also took out the state of Mississippi from the analysis because the state had an unusual WIC market structure over Newsy's timeframe. Prior to 2021, Mississippi was the only state that utilized food distribution centers instead of retail grocery stores to distribute WIC foods to program participants. Therefore, Mississippi state data was inconsistent with the rest of the United States. In 2021, Mississippi transitioned to retail grocery stores for WIC product distribution. Reporters used data obtained from the state of Washington for the fiscal year 2020 because federal USDA Washington state data was incomplete.

The two big things Newsy reporters had to do to the data to make the analysis possible were deduplication and mapping stores to counties. (Note: the Washington state data for 2020 had no apparent duplicates and included vendor county as part of its address information, so was not part of the processing below.)

Duplicates

When Newsy got the data, reporters realized that it contained vendor duplicates. One kind of duplicate was easily explained: The same store could be registered with more than one agency in any given year. For example: In 2020, Shaw's Supermarket at 10 Benning St., Suite 1 in West Lebanon, New Hampshire was registered both with the state of New Hampshire's WIC program and the state of Vermont's WIC program — thus, it was getting counted twice. Newsy saw this frequently with stores registered both with their state WIC, and with one or more tribal WIC programs and areas close to state borders.

The other kind of duplicate was harder to explain. For example, the store at 406 East Main St. in Salem, Illinois was registered twice under Illinois' WIC program in 2011 — once under the name "Salem IGA Foodliner Plus" and once under the name "Salem IGA Foodliner." Sometimes, like with Casey's at 430 College St. in Cedartown, Georgia, a store is registered more than once with the same program under the exact same name and address.

Newsy reporters found that a likely explanation for many duplicates like this is that in a single year, the store may have left and reentered the program. Stores can leave the program for many reasons, including when they lapse in their requirements to continue to participate, or when they change ownership, or restructure in a way that requires them to change their registration with WIC.

These duplicates presented a challenge for Newsy's analysis. If the only WIC store in your county changes ownership three times in a year, reregistering for the WIC program every time, there's still only one WIC store in your county. But the USDA data would count this as three separate stores for that year. If the only WIC store in your county is also registered as a WIC store for residents of a neighboring state, there's still only one WIC store in your county. But the USDA data would count this as two separate stores for that year.

To take out duplicates, Newsy reporters created a standardized street address field. These were transposed to all-capital letters, and reporters did some things to align different ways of specifying the same address, including:

- Standardizing ZIP codes to 5-digit ZIP codes
- Trying to make sure street numbers don't appear twice (sometimes they are included in Street Name)
- Removing suffixes like "ST," "WY," "AVENUE," "BLVD"
- Translating words like "NORTH" to "N" and "HIGHWAY" to "HWY"

Newsy reporters did a spot check of matched stores to confirm Newsy methods were finding true duplicates and adjusted Newsy's standardization methods accordingly. Reporters also looked for stores with the same business name containing a store number (e.g. Ralph's Market #2) to detect more examples of duplicate addresses. Stores with the same standardized address within a year were considered duplicates and all but one was thrown out. This eliminated about 2% of vendors in 2011 and in 2020.

Counties

Newsy decided to analyze access on the county level, which was unfortunately not part of the address data reporters obtained from USDA. In order to determine the county each vendor was in, reporters used a layered approach.

1. Newsy reporters downloaded the [ZCTA-to-county crosswalk](#) from the U.S. Census website and joined it to Newsy's data on ZIP code. If a vendor's ZIP code was wholly contained inside a county, Newsy assumed the vendor was also inside that county. This captured 75% of unique vendors in 2011 and 2020 combined.
2. Newsy reporters flagged a small number of the remaining vendors as difficult to geocode (often due to unusually structured or poorly formed street addresses).
3. Newsy reporters geocoded the remaining vendors by their standardized address, and compared them to a census shapefile of counties to determine what county the vendor was in.
4. Vendors that were not geocoded to a county predicted by the ZCTA crosswalk were combined with the vendors Newsy reporters determined too difficult to geocode. Newsy used the census shapefiles for "place" (incorporated city, town, or equivalent) to determine which places were in which counties. If any of these remaining vendors was in a city, town or place that was wholly contained within a county, reporters assumed that vendor was also inside that county.
5. Vendors that reporters could not assign to a county by any of the above steps, or that were assigned to more than one county, or that were assigned to a county outside of those predicted by the ZCTA crosswalk, were flagged for manual location. Newsy reporters went through these 100 remaining vendors and kept track of where Newsy reporters found them.

The hierarchy of methods for assigning a county to a vendor was as follows: by ZIP code (75% of vendors in 2011 and 2020) -> by geocode (24%) -> by "place" (0.4%) -> manual lookup (0.2%).
Byline: [Rosie Cima](#) and [Rachel Gold](#)