

[This Photo](https://unu.edu/publications/articles/taking-hunger-out-of-equation.html) by Unknown Author is licensed under [CC BY-SA-NC](https://creativecommons.org/licenses/by-nc-sa/3.0/)

|  |  |
| --- | --- |
| Can Grocery Delivery Services Reach Food Deserts? | |
| Fruit bowl |  |
| Rice Data Visualization BootcampJanuary 11, 2020Raquel McLellan |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | |  | |
| The Project | | | | |
| As part of the class ETL project for the week, I chose to investigate whether grocery delivery service may be a way of reaching areas known as food deserts.  In choosing the sources, I was able to locate the Food Access Research Atlas from the USDA ERS division, which is a main source in tracking the variables that define Food Deserts for people across the US. The next source considered was a well-known grocery delivery service, Instacart to see if their service areas may include these low access/low income areas of concern. While there are other services and potentially even availability from some grocery stores, this is national entity so it seemed that finding information on a broad basis to compare to the USDA ERS data would be more practical.  However, due to the datasets and what they contained, I found that I had to pull additional data sources including the states and their abbreviations from Wikipedia and a third-party US zip code source as it contained the County data by City/State combination. | | | |
| Fruit bowl | |  | | |

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
| In walking through the data there were instances of many cities or counties multiplied throughout, leading to larger data to manipulate and less accuracy in the results. It would have been better if I could have either used zips or matched up the census tracts, but the availability of that information and connecting to my main sources was not readily apparent.  In trying to normalize the data, I removed what duplicates that I could. However, I did have to go into the csv and remove the words: County, Parish and Borough from the County info on the zip code file due to it not matching with the USDA County descriptions. Even with that there were still other issues such as the spelling with or without spacing on the names of cities. It was some of this inconsistency that led me to using MongoDB rather than PostGres as I was not certain I could create the required relationships connections, even though ideally I would want to for an analysis like this if I had more time.  In walking through the data there were instances of many cities or counties multiplied throughout, leading to larger data to manipulate and less accuracy in the results. It would have been better if I could have either used zips or matched up the census tracts, but the availability of that information and connecting to my main sources was not readily apparent.  In trying to normalize the data, I removed what duplicates that I could. However, I did have to go into the csv and remove the words: County, Parish and Borough from the County info on the zip code file due to it not matching with the USDA County descriptions. Even with that there were still other issues such as the spelling with or without spacing on the names of cities. It was some of this inconsistency that led me to using MongoDB rather than PostGres as I was not certain I could create the required relationships connections, even though ideally I would want to for an analysis like this if I had more time.  The ETL performed on the data beyond the manual changes to the Zip Code csv included:   * Using Splinter to scrape the Instacart service region page to obtain the cities they are active in and storing that in MongoDB as well as creating a reference csv file for other datasets of the states services; * Isolating the key data variables from the large csv file from the USDA ERS site and limiting it to the states in which Instacart provides service and storing in a collection; * Using Pandas to scrape the Wikipedia page that held the States and their abbreviations and capturing it in a CSV file; and * Joining the appropriate files to the County and State Abbreviations as necessary to allow a final connection between the Instacart cities and the identified food deserts in the states Instacart operates today.   While I was able to reduce and clean my files to make them more manageable, there are likely ways to reduce some of the steps taken:   * potentially using maps, * string syntax to remove the items I did manually, * less steps to connect the files to get to the final collection, * and leverage SQL more for the joins likely could have been more efficient as well.   Fruit bowl  It is likely debatable if the code is maintainable. Initially, the notebooks were being written with a lot of detail on the steps and intent of the code. However, as the time for the project to be completed approached, the code took precedence over the explanation so later notebooks may not be as easily understood. I did find during the course of the project that I had to readjust the code on the splinter web scrape when the site updated and that was easily accomplished.  MongoDB Schema  The following is the illustration of the database and collections within it: | | | |
| |  |  | | --- | --- | | Instacart\_Cities | | | Id | **City/State** | | | |  |  |  | | --- | --- | --- | | Instacart\_Counties | | | | Id | **City/State** | **County** | | |
| |  |  |  |  | | --- | --- | --- | --- | | Zipcode\_CouNTY\_Data | | | | | Id | **City-State/County** | **City-State** | **County** | | | | |
| |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | food\_deserts | | | |  | |  | | Id | **CensusTract** | **State Abr** | **County** | **Urban** | **Food\_Desert** | | | | | |
| |  |  |  |  | | --- | --- | --- | --- | | Food\_Deserts\_With\_SVC | | | | | Id | **County** | **City/State** | **Food\_Desert** | | |  | |
| Fruit bowl | |  | |
|  | Potential Use Cases:  * Searchable access to database to find areas that can be serviced for policy review and philanthropic investment to fight hunger. * Using the data to heatmap areas of the US where food insecurity is a meaningful social determinant of health and working with employers to see if offering a benefit such as Instacart to employees helps provide access to food. * Identify areas with higher factors of food access issues (higher food desert score) and see which cities are more impacted. | |  |
|  | Fruit bowl | |  |