# BACKGROUND

Food deserts (“FDs”) are areas where people have limited access to a variety of affordable foods. Typically, these regions include low income areas, high levels of unemployment, limited transportation, large or sparse populations, or limited food retailers that provide food options[[1]](#footnote-1).

# ISSUE

Group 10 team members Hardik Vadel, Igor Pavlunin, Jana Warri, Katy Fuentes, and Oluwadamilare Alade reviewed 2015 US Department of Agriculture’s ERS (Economic Research Service) FD data1 Gazetteer Files linked to the 2010 Census[[2]](#footnote-2) and Centers for Medicare and Medicaid Services[[3]](#footnote-3) public information sources to analyze different social and economic implications of FDs.

The ERS Files include over 148 variables of which the team focused on the following:

LILATracts\_1And0, LILATracts\_halfAnd10, LILATracts\_1And20- Low Income Low Access (“LILA”) tract at 0.5 mile for urban areas and up to 20 miles for rural areas. Each of these were flagged with a 0 and 1 (0 = No FD and 1 = Yes FD) and when combined could amount to a Food Desert Score of 1 through 3, with 3 being the highest/worst.

HUNVFlag - Tract with Housing Units without Vehicle (“HUNV”)

PovertyRate - Tract poverty rate.

MedianFamilyIncome - Tract median family income.

TractSNAP - Tract housing units receiving SNAP benefits.

The Gazetteer Files include geographic information linking each census tract to its latitude and longitude.

The Centers for Medicare and Medicaid Services information includes actual spending per capita on 21 chronic conditions.

# ANALYSIS[[4]](#footnote-4)

***Families on SNAP benefits & Geospatial Visual (Hardik)***

Food desert information per census tract was linked to Gazetteer files with geospatial data. This provided a data frame with all data and geographic locations. An API callout to Google maps was written to map FDs over a United Sates map. A color gradient was applied with red indicating a severe FD score of 3 and yellow as FD score of 1. Based on the data its evident that low income low access populations are more concentrated on the two coasts and in the Southwest portion of the country. A more granular plotting could be undertaken since there were roughly 74,000 data points available across the country but due to limitations of time this was not undertaken.

Mapping the correlations of households receiving SNAP benefits to a FD score involved first normalizing the SNAP data across number of households. This was done by dividing households receiving SNAP benefits by total number of tract households. This percentage information was appended to the data frame as “SNAPNorm”. A few discrepancies between the Gazetteer files and the Census data were altered in the analysis as follows:

1. 25 Census tracts which did not have latitude and longitude information were removed.
2. 2,000 of the total 74,000 Census tracts that showed SNAP benefits households higher than total households and those with zero households were removed.
3. To streamline the analysis without investing too much time in data cleansing, an assumption was made, that there is a one to one relationship between household count and SNAP count.

***Housing Units without Vehicle in food deserts (Igor)***

HUNV data exploration and correlation starting from the merged 2010 Census and Gazetteer files to enable latitude longitude geolocations in order to plot FDs areas over the Google Map. First step was to normalize tract population and tract housing counts versus total population and total housing units and drop tracts with zero population and/or housing units. The focus area consisted of housing units without a vehicle (HUNV) and low income population (LOWI). Data normalization was performed by dividing “per tract” values to “total” values and multiplying by 1,000 to get normalized values close to real scale. Next step, was to summarize tract data for each state and plot top five states with highest HUNV numbers for FDs 1 and FDs 3. Testing HUNV data for mean distribution revealed that HUNV (and LOWI) data are not normally distributed but following Poisson distribution model instead. This fact dictated that LinRegress and TTests will not provide reliable correlation outputs, since both tests have normal data distributions as a key requirement. In order to explore any possible correlation between HUNV and LOWI per different FD score areas a plot chart was utilized to summarize HUNV and LOWI data per state. Lastly, a LinRegress and two TTests was utilized for demonstration purposes, which also provided a visual correlation as the prior model indicated.

***Chronic condition spending in food deserts (Katy)***

2015 actual per capita spending by Medicare and Medicaid on 21 chronic conditions was merged with the Census information by state and county because the Census information was based on an individual county tract level. Although data exists for 21 chronic conditions analysis was only performed on Depression, Diabetes, Cancer spending relative to FD categories 1 and 3. There were several data ingestion issues in merging the two data sets because there were multiple extra spaces and other special characters in the source data.

Other data analysis performed includes computing nationwide totals based on overall spending by chronic condition and state provided statistical data relative to total, median, mean, and minimum amounts spent by condition. One observation is that FD category 3 spending on Drug Abuse/Substance Abuse at the state and county level provided median spending of $141,926.80. This analysis helped identify Virginia as the state with the highest population of category 3 FDs with the most spent on Stroke conditions. Lastly, total “National” spending across the 21 chronic conditions is generally evenly distributed between 3% through 6% per condition.

***Poverty rates in food deserts (Jana)***

Grouped data by state to narrow the area and calculated the median poverty rate % to use in my analysis. In addition, limited my population to area with a food desert core of 3 or more and referencing income. In 2010, the median house hold income was $49,445 with 58,797 scored areas for food deserts. Of those, 14,067 designated area had a food desertes rating 3 or higher. Meaning 23.92% of the US lived in a designaed moderate food desert. The Census data was limited with no tracking for household size, age of age of ocupants.

P-value of .257562 indicates there is only 25% probabliy the high poverty rates in moderate food deserts is a random occurrence. I also calculated the mean proverty rate percentage and standard deviation of the data set and determed the limtes to detmine the data is normally distributed.

***Median Income and food deserts (Oluwadamilare)***

# CONCLUSION

***Families on SNAP benefits (Geospatial Visual) (Hardik)***

Based on state specific data analyzed for three most populous states, there is a linear correlation between number of households receiving SNAP benefits and intensity of FD impact. It appears that nearly 83 million US citizens or 27% of the 2015 total population of were impacted by FD. It is also evident from the table below that the higher the FD score, the higher the percentage of SNAP benefits usage.

National Data – SNAP Benefits Usage and Food Desert Score

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CATEGORY** | **VALUES** | | | |
| FOOD DESERT SCORE | 0 | 1 | 2 | 3 |
| Mean of Households with SNAP (%) | 10.30% | 24.01% | 16.84% | 22.705% |

Along with these insights, it was reported that the US spent $70 billion on SNAP benefits. A more cost effective remedial policy could be tax incentives to grocery store chains to setup in FD locations in some of these regions. However, further cost benefit analysis towards this should be conducted.

***Housing Units without Vehicle in food deserts (Igor)***

Plotting the top five states showed that HUNV distribution is not even or equal in different FD scores. FDs 1 has a higher values of HUNV in a random state when compared to FDs 3 HUNV values. As such, it appears that HUNV have a different distribution in different FD scores.

Plotting manual visual correlations revealed following findings:

FDs 1 has the highest values of HUNV across the states

FDs 2 has the smallest values of HUNV across the states

FDs 3 has higher HUNV values than FDs 2 but less than FDs 1

To address highest impact of HUNV government organizations should focus on FDs 1 areas primarily. Some possible support options include but are not limited to short/long term low interest loans, tax breaks for auto-dealers to allow lower prices, or attract businesses to raise employment rates.

Low income (“LOWI”) populations correlate almost linearly with HUNV in FDs 1, FDs 2 and FDs 3 leading to conclusion that FDs 1 states (or areas) have the highest impact in terms of HUNVs and LOWIs.

HUNV and LOWI impact per FD scores in US includes:

FDs 1 HUNV count – 1.720 million houses

FDs 2 HUNV count – 0.068 million houses

FDs 3 HUNV count – 0.890 million houses

FDs 1 LOWI count – 6.000 million people

FDs 2 LOWI count – 0.520 million people

FDs 3 LOWI count – 4.100 million people

***Chronic condition spending in food deserts (Katy)***

While the states vary, spending on Depression consistently ranges between $18,000 through $23,000 in both FD categories 1 and 3. Per capita spending on Cancer and Diabetes are both substantially higher in FD category 3 when compared to FD category 1. The Cancer and Diabetes data also suggests that there is a stronger positive correlation between per capita spending and higher populations in category 3 FDs than category 1 FDs.

***Poverty rates in food deserts (Jana)***

The upper limits of the data has an average poverty rate or 32.44% the lower limits have a 15.96%. In addition, the income followed similar trends. Mississip, Alabama, Louisiana, Tennisee and New Mexico had the highest rates with an average income of $38,000. In contrast, Maryland, Connectucut, Massechusets, New Hampture and New Jersy have the lowest rates with an average income of $59,000. In conclusion, a change in income of 20-30K is presumably enough to move a house hold from a moderate food desert area .

***Median Income and food deserts (Oluwadamilare)***

1. <https://www.ers.usda.gov/webdocs/publications/45014/30940_err140.pdf> [↑](#footnote-ref-1)
2. <https://www.census.gov/data/developers/updates/new-discovery-tool.html>

   <https://www.ers.usda.gov/data-products/food-access-research-atlas/>

   <https://www.ers.usda.gov/data-products/food-access-research-atlas/download-the-data/>

   <https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.2015.html> [↑](#footnote-ref-2)
3. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/CC_Main>

   [Spending County Level: All Beneficiaries, 2007-2017 (ZIP)](https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/Downloads/CC_Spend_County.zip)

   4 Data and analysis files are located on GitHub repository: <https://github.com/hardik77us/ProjectGroup10_1> [↑](#footnote-ref-3)
4. [↑](#footnote-ref-4)