***Healthy dollar store models and community demographics***

***Research proposal***

PI: Dr. Jerry Shannon, University of Georgia

Co-reseacher: Maya Rao, University of Georgia

**Overview**

The purpose of this research is to identify where two major dollar store chains, Dollar General and Family Dollar, offer store models with enhanced access to fresh produce. For Dollar General, this includes both Dollar General Market and retailers identified on the corporate website as providing fresh produce. For Family Dollar, it is only the latter. This research builds on existing research examining dollar store locational strategies and their relationship to neighborhood sociodemographic characteristics and the broader landscape of food retail.

**Research Questions**

1. Is there a significant geographic pattern in where fresh produce is available for each of these two chains? If so, what is the nature of that pattern?
2. What are the demographic characteristics of census tracts with higher than average proximity to dollar retailers offering fresh produce? How do these differ from tracts with lower than average proximity to these retailers?
   1. Sociodemographic traits include household income, poverty rate, racial classification, car ownership, and urban/rural status.
3. How does the proximity of healthier dollar retailers models for a census tract compare to the proximity of supermarkets?
4. What are the potential implications of this research for efforts to ensure equitable access to healthy food in low-income communitities?

**Data**

Retailer listings would be obtained from Safegraph, a POI provider who also includes information on store traffic patterns based on cell phone data. Dollar General Market retailers would be identified via name from these data. This would be matched with listings from the two retailers’ websites, including whether retail locations offer fresh produce. Supermarket listings would also be obtained through Safegraph.

Demographic information would be obtained from the U.S. Census American Community Survey, downloaded through the Census API.

**Methods**

To calculate store proximity, we will use the Mapbox Directions API to identify drive times from census tract centroids to the closest and third closest dollar retailers from each chain as well as the same two figures for retailers that also feature fresh produce. We will obtain these numbers by first identifying the five retailers with closest Euclidean proximity (straight-line distance) via computational methods and then querying the travel time to each retailer using the API. This is a cost effective approach that has proved effective in past research.

We will use these two figures to set up a proximity ratio (PR):

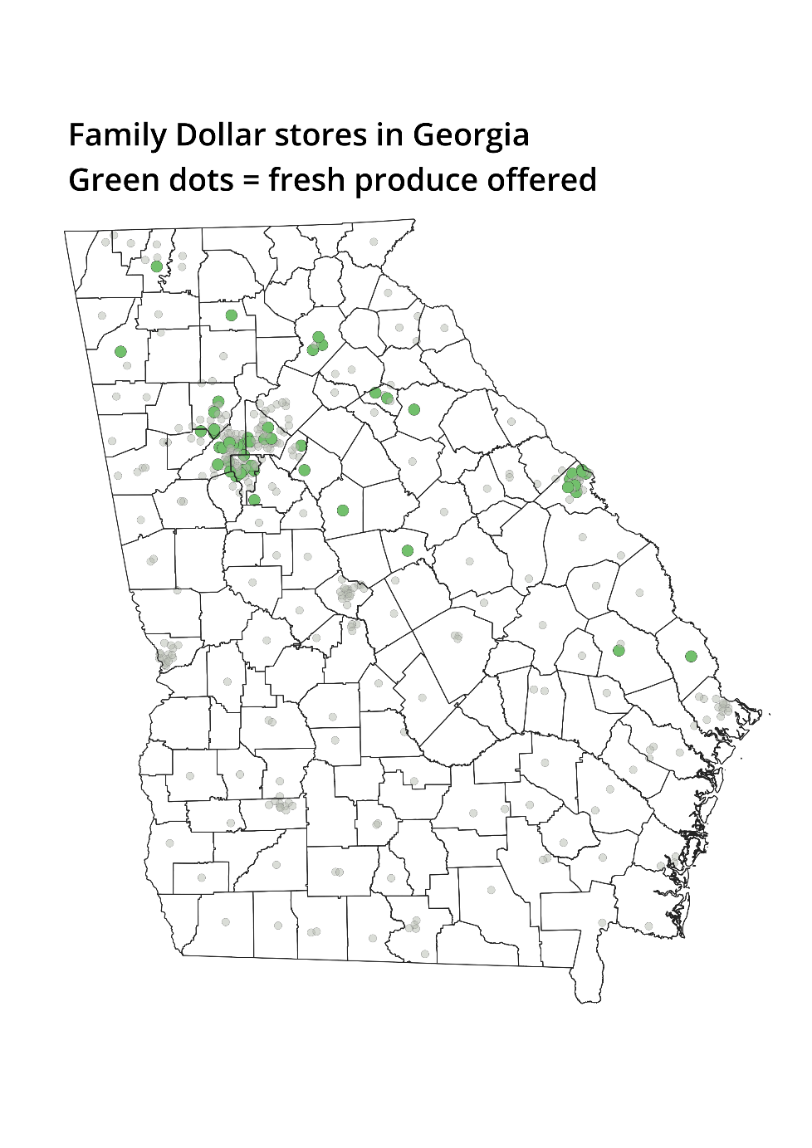
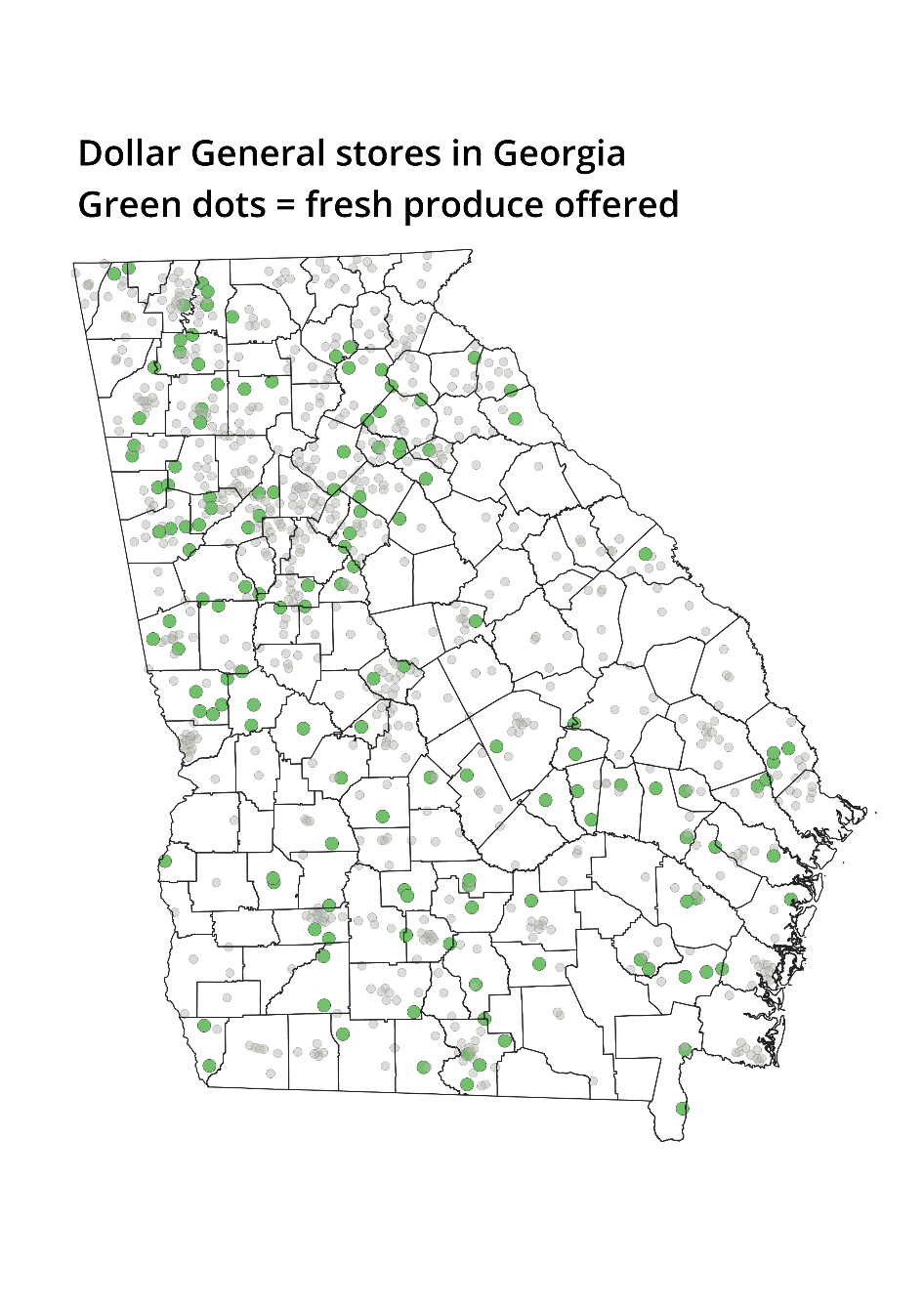
Here, PR is the proximity ratio, dtp is the drive time to a retailer from a given chain including fresh produce, and dta is the drive time to any retailer in the chain. This number will approach 1 if all retailers offer fresh produce, but a higher figure will indicate relatively greater distances to retailers with these items. We use both the closest and third closest retailer to assess density as well as basic proximity.

We will use exploratory analysis to assess patterns in the PR, creating maps to visualize the overall pattern and using hot spot analysis to highlight those that appear significant. We will also create a regression model with the PR as the dependent variable and the demographic data as the independent variables to see what factors are associated with greater relative access. Additional variables may also be included after exploratory analysis if warranted.

To determine the relationship of dollar store produce availability to supermarkets, an additional regression model will be created with just population density and supermarket driving time as independent variables and PR as the dependent variable.

**Preliminary findings**

The most difficult aspect of this analysis is obtaining the list of retailers offering fresh produce. To test the viability of this approach, we use the RSelenium package in R to scrape the amenities offered by each retail location on the corporate websites. Maps of all stores are shown below, with stores offering produce shown in green. Dollar General Market locations are not included but will be in the final analysis.



**Timeline**

Start date: 5/15/23

Data collection complete: 6/15/23

Analysis complete by 7/15/23

Final report submission: 10/1/23

**Costs**

Safegraph

* **$150** for one month subscription to download data

Mapbox Directions API (for drive times):

* 85,000 census tracts \* 5 queries per retailer type \* 5 retailer types (DT/FD with/without produce + supermarkets) = 2,125,000 queries. ([pricing guide](https://www.mapbox.com/pricing))
* **$3,000** total for these queries

Remote server for data collection

* Three month subscription to [Posit Cloud](https://posit.cloud/plans): $75 \* 3 = **$225**

Consultant payroll

* Dr. Jerry Shannon (~40 hours @ $175/hour): **$7,000**
* Maya Rao (~50 hours @ $60/hr): **$3,000**

Total funding requested: **$13,375**