

# Distribution of Main Grocery Stores in Wisconsin

Juan Berón

August 2020

## 1 Introduction

It is well documented that certain stores are located at areas with low population density or non-urban. However, the location of a new store, for example, a WalMart super center, depends on many different factors besides only the population density. Factors such as the presence of competitors, the closeness to another store of the same brand and the population density are interesting variables that could determine if a retail store should open or not a new location.

Besides the well-known examples such as WalMart, it is interesting to explore the competitors in the category of grocery stores. This work will focus only in Wisconsin, however, it is desired to expand to other areas of USA. This work will focus mainly on the identification of locations with low population that have a central point. Then, the grocery store options that this places have access to will be explored.

## 2 Data Description

The main sources of information are the census of USA and the data from foursquare. The United States Census provides population and location data in the following link:

<https://www.census.gov/data/tables/time-series/demo/popest/2010s-total-cities-and-towns.html#tables>

Figure 1a shows the first five rows of the table provided by the census about Wisconsin. From this, the information that will be used is the County, Name and Population in 2010. This information is enough to get the latitude and longitude at wikipedia.org as indicated in Figure 1b. The latitude and longitude are given to foursquare to obtain the Grocery stores around the locations.

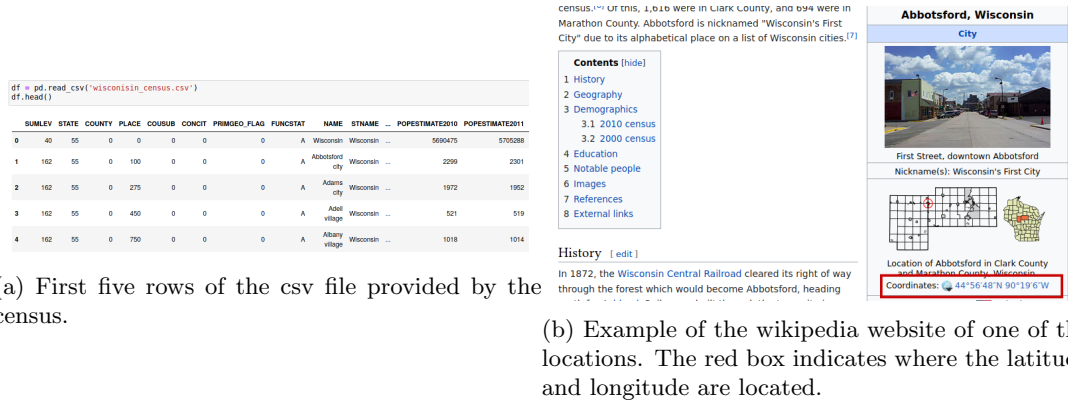


Figure 1

### 3 Methodology

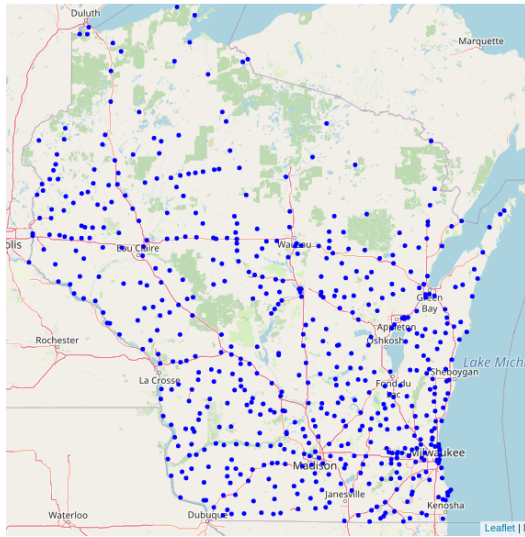
#### 3.1 Cleaning the Data

1. Dropping entries that were not formed in 2010 which are marked with an 'A' in the census population
2. Dropping double entries that were necessary for administrative purposes and were identified by having the county number 0
3. It was identified that Wisconsin had four types of locations included in the census: county, city, village and town. The type that was related to county was drop since our study will focus on cities like location. Additionally, town type was dropped because it lack a central are and it is made of places that were not included in cities or village in a particular county.
4. Merge locations whose population was divided because being located in the edge of two or more counties. The name of the location is the same, but there are two or more entries with the same name and different county number.
  - (a) Before merging this place, it was checked if the location was in fact the same city divided by borders. To do this, information about the cities was scrapped from wikipedia.
  - (b) Each place identified as being divided was confirmed by wikipedia information. Then, the entries were merged by adding the corresponding populations and keeping the county with the highest population.
5. The latitude and longitude of the cities located divided by counties divisions were gathered from wikipedia. For the other locations, the library *geopy* was used. An address made from the name of the place, the county name and the abbreviation of wisconsin was given to geopy to find the latitude and longitude. The places that geopy did not find any information were added manually.
6. The analysis was limited to villages and cities of four, third and second class, i.e., locations with less than 39000 people.

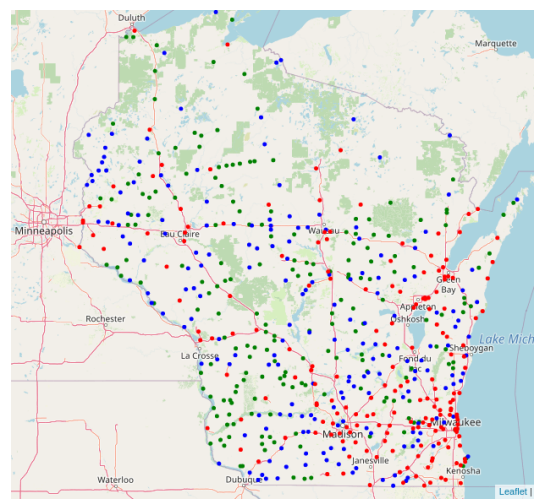
7. Latitude and longitude values were confirmed visually by plotting their values in a scatter plot, the places that had distinctive different values were fixed.
8. Each location was marked in the map of Wisconsin with the help of *folium* to make a final visual confirmation of the latitude and longitude values.
9. To use foursquare, it is necessary to provide a radius for the search query. To avoid intersection I created a matrix with the distance between each of the locations. Then, the minimum distance between each location and a different locations was recorded in a vector. The median of that vector divided by two was the chosen radius.
10. Foursquare search queries were issued for each location. The search was limited to the category Food & Drink Shop which included different types of food services, but in particular it included Grocery Store and Supermarket.
11. Foursquare returned desired location information, but it included repeated entries. For example in Adams city A-F County Market was repeated twice, however, Adams city only has one A-F County Market. Therefore, the approach was change to if the location has or does not have a particular food provider. Then, each repeated food provider name per location was erased.
12. Further study was going to be according to clusters of locations. The clusters were selected according to the population and for that kmeans was used and 12 clusters were made.
13. I decided to narrow the scope and study the two most common brand of food providers: Piggly Wiggly and Walmart. The number of location that had this food providers were count according to the selected clusters.

## 4 Results

The maps for visual validation of the values for latitude and longitude are depicted in Figure 2a. A distribution of the population is presented in Figure 2b. Table 3 presents the counts of the two most common food providers in Wisconsin according to the clusters given with the kmeans method. Finally, Figure 4 gives information of how the food providers are distributed across the population.



(a) Places included in the study showing a visual validation of the values of latitude and longitude



(b) Population distribution. Green: between 61 and 757 people. Blue: between 762 and 2580 people. Red: between 2584 and 37920 people.

Figure 2

	count	Piggly Wiggly	Walmart	Piggly Wiggly P	Walmart P
mids					
584	121.0	12	4	0.099174	0.033058
1664	86.0	9	8	0.104651	0.093023
3020	63.0	18	6	0.285714	0.095238
4568	30.0	11	10	0.366667	0.333333
6408	20.0	8	11	0.400000	0.550000
8597	26.0	9	19	0.346154	0.730769
11478	24.0	10	10	0.416667	0.416667
14713	15.0	6	9	0.400000	0.600000
19060	11.0	3	5	0.272727	0.454545
24664	8.0	6	6	0.750000	0.750000
28898	4.0	1	3	0.250000	0.750000
35828	7.0	5	4	0.714286	0.571429

Figure 3: Mids is the middle value of the range of the cluster. Counts is the number of locations in this cluster. Piggly Wiggly column indicates the number of locations in the cluster with a Piggly Wiggly. Walmart indicates the number of locations in the cluster with a Walmart. The last two columns is the number of locations with either a Piggly Wiggly or a Walmart over the total count of locations in the cluster

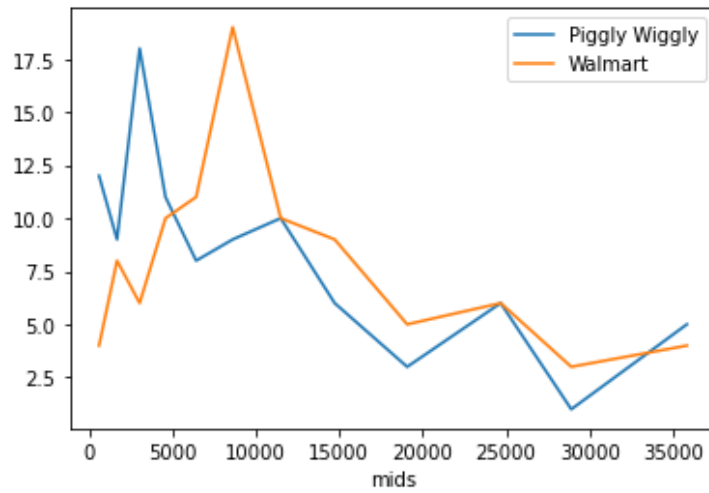


Figure 4: Distribution of the number of location with the food provider according to the population in the location.

## 5 Discussion and Conclusions

Figure 4 is presents the final outcome of the project where we see the that Walmart allocates its stores in locations with middle population while Piggly Wiggly is more often found in smaller locations with less population. Nevertheless, it is interesting to know what type of Walmart is being constructed in each location since they vary in size. Figure 2b shows that there is a clear difference in population density between the south and north Wisconsin. Therefore it would be interesting to know how is the distribution of this main food providers according to the latitude and longitude. This project allows to look at possible variables that help to determine where the next Walmart or Piggly Wiggly will be.