

# Testing the Spatial Economics Theory with Different Industries and Different Cities

Jesse Russell

June 15, 2020

## 1. Introduction

It is a long-held tenet of economics that businesses often have a rational incentive to locate nearby their competitors, rather than locate farther away from their competitors. Harold Hotelling (1929) outlined the rationale nearly a hundred years ago and empirical research over the years has found this idea to hold true in many sectors. This idea of "spatial economics" considers how businesses can increase the segment of the market that is closest to them by positioning themselves in the center of the geographic space. Each business has the same incentive, so they all tend to locate in the same central area.

The question I propose to examine in this project is whether different kinds of businesses (as labeled in the Foursquare data) cluster more closely than others. For

example, do coffee shops locate nearer to each other in a given city than do Italian restaurants. Or do Italian restaurants cluster more closely in that city but not in another? I propose to look at a number of different kinds of venues and calculate a mean distance measure for each. This will include calculating the center position of all the identified venues and then calculate the average distance from the venues to that center point.

Knowing the center of each kind of business cluster will allow me to calculate the geographic distance of each business to the center and to then calculate measures of central tendency and dispersion. These measures will allow me to make comparisons across industries and across cities.

### 1.1 Background

Both from an academic view (why do businesses locate where they do) and from a practical view (where should I locate my business) there is a puzzle of whether it is better to locate a business near its competitors or far from its competitors. At first blush, it might seem intuitive to locate far from competitors so that you can divide the geography and be closer to some customers who live farther away from the

competitor. But, if there is an assumption that, everything else equal, people to go whichever business is closest, then a business should locate close to its competitor. The idea here is that a business can cut the biggest chunk of geography for itself if it cuts from the center. Moving outward away from the competitor gives the center away to the competitor.

## 1.2 Problem

The problem is that some kinds of businesses in some kinds of geographies might not follow this theory. There might be some kinds of business where placing a business at the center is not beneficial. For example, businesses that people tend to walk to cannot locate themselves far away. Or for communities that rely more on public transit, there might be an incentive to locate closer to the community to make it easier for them to get to.

There are also some kinds of businesses that tend to go with some kinds of communities more than with others. For example, laundry businesses (such as laundromats) general serve neighborhoods where there are a lot of apartment buildings and where people do not have washers and dryers at home. These neighborhoods tend to be lower income where things like

housing type and transportation options play a big role. In contrast, banks and financial services do not tend to locate in lower income neighborhoods. Banks and financial services businesses are generally more likely to locate in business districts and higher income neighborhoods.

There is also the dimension of how different cities are laid out. Historical trends in housing, and even housing discrimination, have left a distinct impression on how the borders of neighborhoods are down, how higher and lower income neighborhoods are placed near and far from each other, and how residential neighborhoods relate to city centers and business districts.

## 1.3 Interest

This question has an academic application, but it also has a very real implication for both a) businesses calculating where they would be best located, and b) how cities can understand how to ensure all neighborhoods in the city have access to services and businesses. A person living in a neighborhood with a density of laundromats might have a different set of opportunities and hurdles than someone living in a neighborhood with a density of banks.

## 2. Data acquisition and cleaning

### 2.1 Data sources

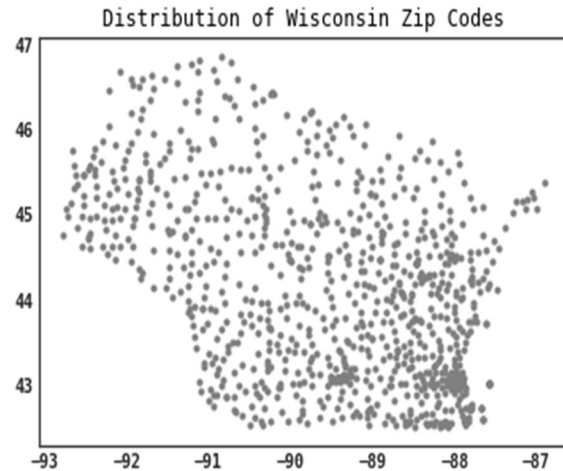
I selected two Midwest states to begin the analysis: Wisconsin and Minnesota. For each state I collected location data, including zip code and latitude and longitude, from the opendatasoft.com public API.

I also collected data on businesses, including latitude and longitude, as well as business name from the Foursquare API. Additionally, I downloaded map information from the Folium Python package.

Information from these three sources were used to address the question about location and density of laundry businesses and banks in Wisconsin and Minnesota.

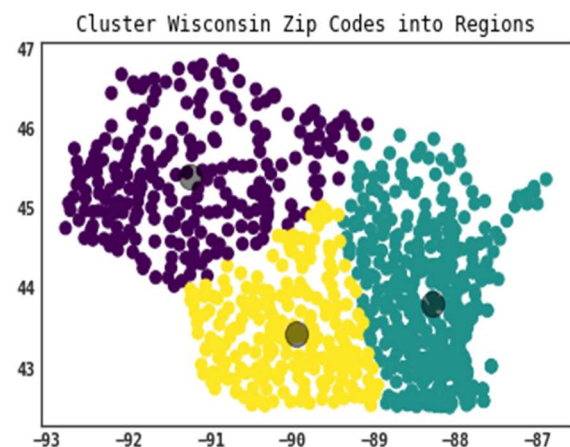
### 2.2 Data cleaning

Data from both API sources were retrieved as json files. The `json_normalize` function of the `pandas.io.json` package was used to flatten the files and convert them into Pandas dataframes. From the Pandas dataframe, I selected the relevant features and renamed columns as necessary. The end result was a dataframe focused on cities, zip codes, latitudes and longitudes, and businesses.



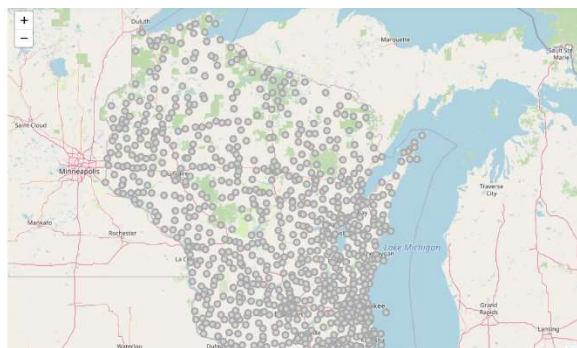
## 3. Clustering by zip code

As a first step of the analysis, I examined the distribution of Wisconsin and Minnesota zip codes across the state. A simple scatter plot shows the distribution of zip codes across the state and shows a distinct area of density around Milwaukee and south to Janesville. There is also a small dense area around Madison and surrounding areas.



### 3.1 Clustering Wisconsin

To see if these three areas are also k-means clusters, I ran the clustering algorithm. Again, it is clear that Milwaukee and southern suburbs make up one cluster, while the Madison area makes up another cluster, and the north west areas, including Eau Claire and La Crosse, make up a third cluster. These clusters are intuitive. The zip codes are not evenly divided, though. The Milwaukee cluster has nearly double the zip codes as the other two clusters.

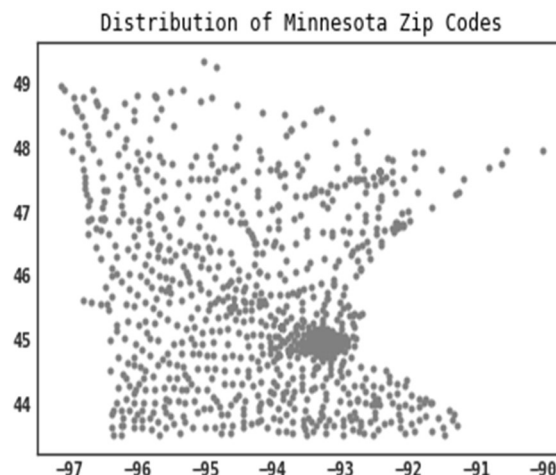


For another look, I initialized a Folium map and plotted Wisconsin zip codes on it. This map reinforces the view that Milwaukee and its southern suburbs and the main urban center, with a cluster around Madison, and a large area of smaller north western cities.

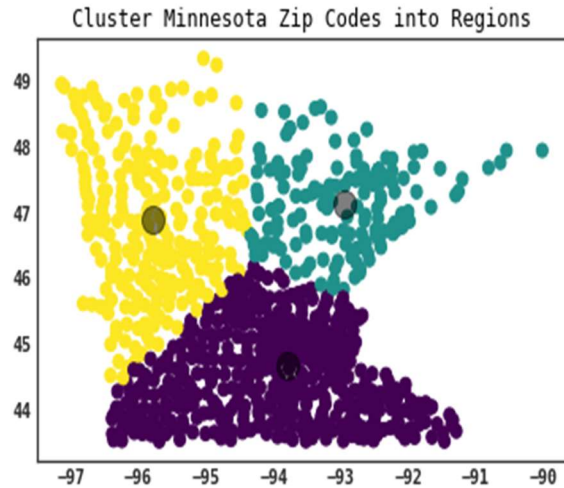
### 3.2 Clustering Minnesota

Similarly, I plotted Minnesota zip codes with a scatter plot. This shows an overwhelming density around Minneapolis

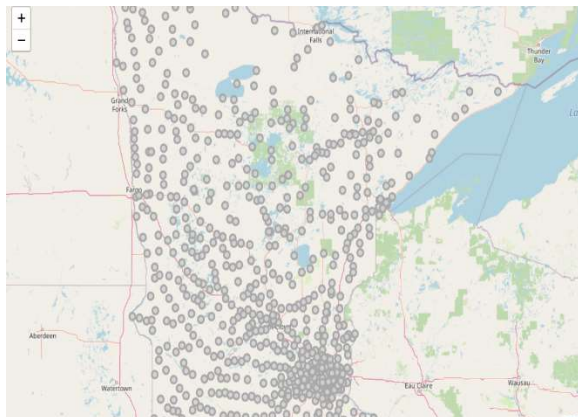
and Saint Paul. There are some suburban areas surrounding the Twin Cities, with very few zip codes to the north and west.



A k-means clustering of zip codes groups nearly all the populous areas into one cluster around Minneapolis/Saint Paul and surrounding suburbs and two other clusters. The Minneapolis/Saint Paul cluster contains more than 60% of all zip codes. A Folium map shows Minneapolis/Saint Paul at the center of a set of concentric circles, dense in the middle and getting progressively sparse moving away from the center.



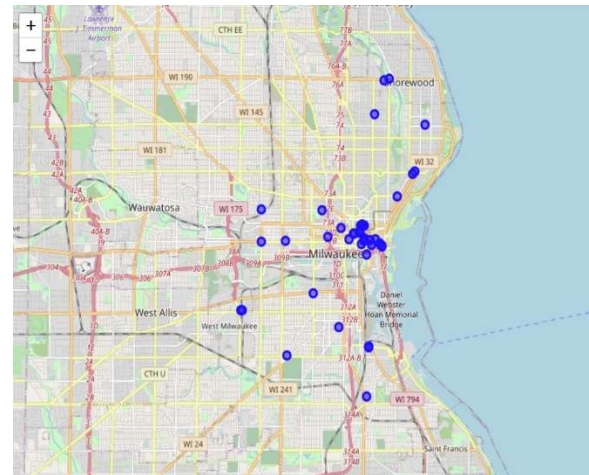
While zip code density is limited in that it can tell is, it is notable that the pattern of density and sparsity is different for the two states and their two metropolitan areas. Minneapolis/Saint Paul sit at the center of a radial pattern of zip code density. It decreases with each step away from the center. Milwaukee, on the other hand, has few suburbs to the north, some suburbs outside of Milwaukee, and the bulk of suburbs heading south (toward Chicago). These differences in geographic organization might have a direct impact on the business density analysis.



## 4. Laundries

For the analysis, I isolated zip codes for Minneapolis/Saint Paul and for Milwaukee. This allow for an analysis based on a single metropolitan area in each state.

As a first step, I plotted all laundries on a Milwaukee map and a Minneapolis map. These maps show something similar and dissimilar. They are similar in that most laundries are near the downtown area. They are not in suburban areas. They are dissimilar in that the Milwaukee laundries are spread more north/south, and the Minneapolis/Saint Paul laundries are spread more circularly around the city center.

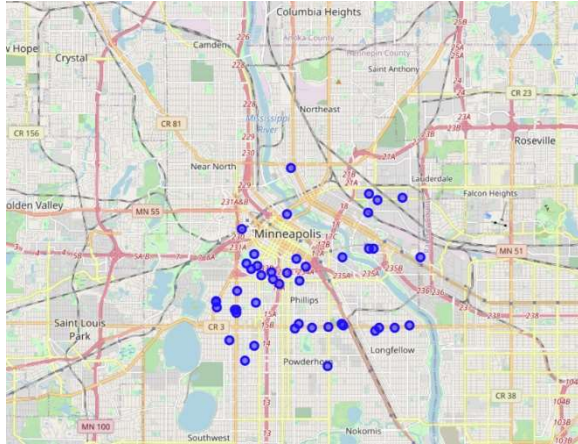


### 4.1 Milwaukee Laundries

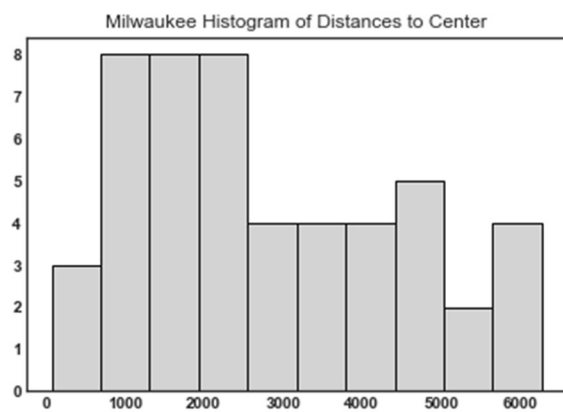
Foursquare identified 49 laundries in Milwaukee. I calculated the geographic center of these laundries at latitude 43.0406 and longitude -87.9112. This is near the city



center of 43.0389, -87.9065. This indicates that laundries are centered on the city center, not on any other part of the metropolitan area.



The calculation used the Haversine formula, which calculates distances over a sphere, like Earth, rather than as distances across a flat surface. This method gives a more accurate measure of geographic distance.

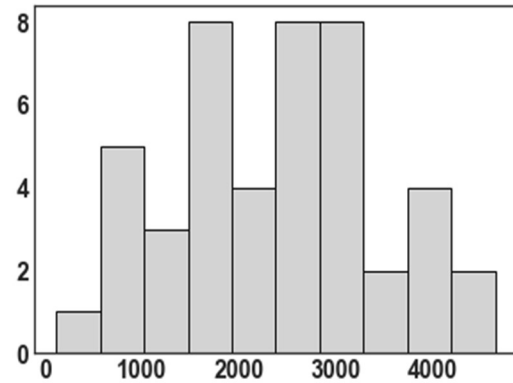


Once the center was identified, I calculated the distance from each Milwaukee laundry to the center. A simple histogram of those distances shows a large

number of distances are short with a tail leading toward longer distances.

A calculation of measures of central tendency and dispersion show that the average distance of a laundry to the center of laundries is 2805 meters. The median is a little shorter, at 2,196 meters, representing a skew in the distribution. The standard deviation is relatively high in comparison to the mean and median at 1,710 meters.

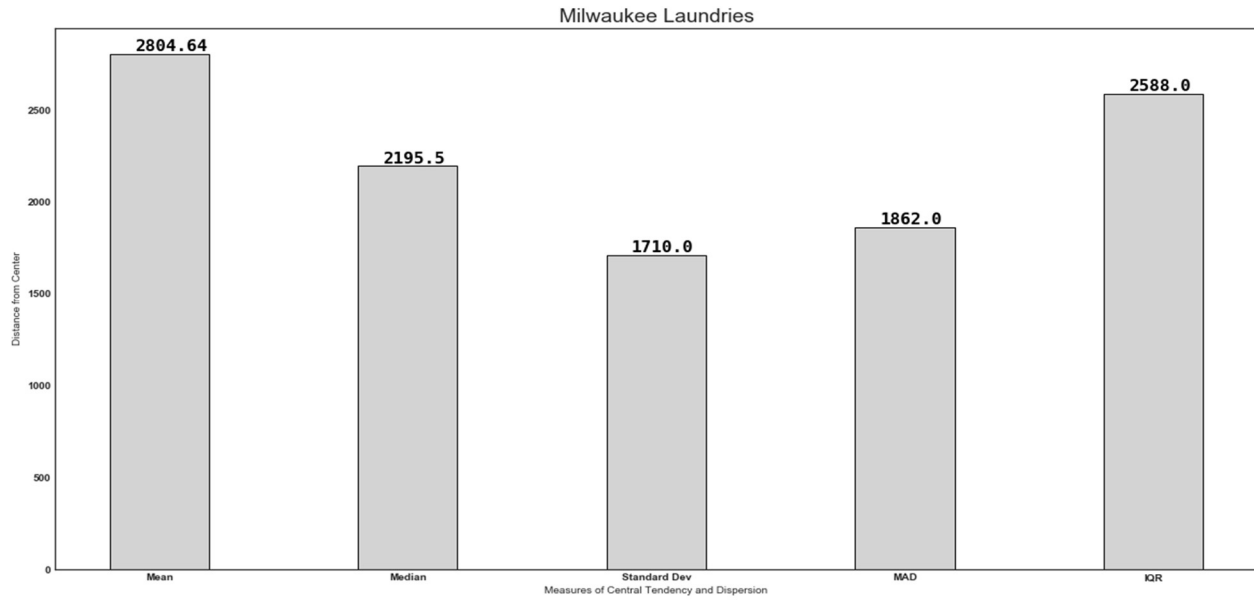
Minnesota Distribution of Distances to Center



Possibly a better indication of variance is the mean absolute deviation, which was a similar 1,862 meters while the inter-quartile range was broader at 2,588 meters.

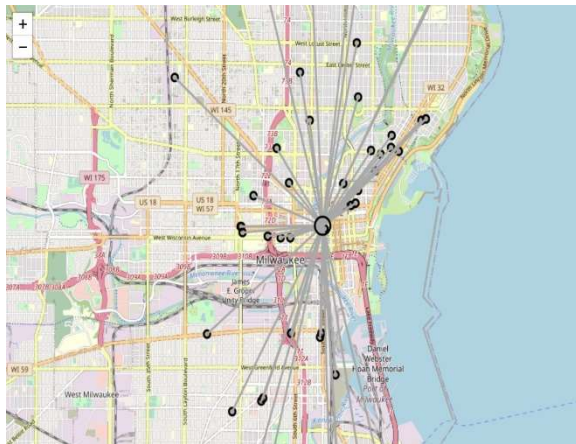
#### 4.2 Minneapolis/Saint Paul Laundries

Foursquare identified 44 laundries in Minneapolis/Saint Paul. I calculated the geographic center of those laundries at latitude 44.9615 and longitude -93.2599. As with Milwaukee, the laundry center was



near the city center of latitude 44.9778 and longitude of -93.2650.

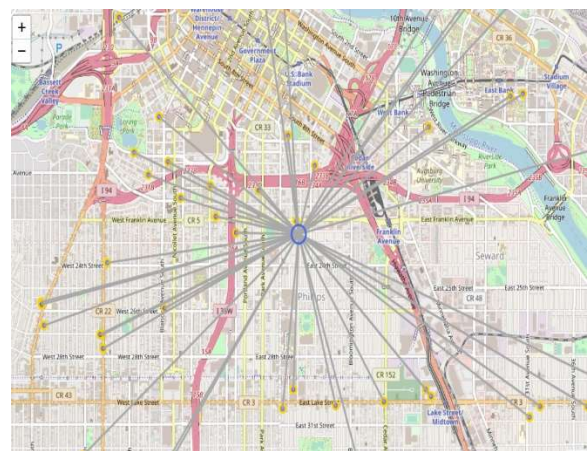
Again, this suggests that laundries are centered on the city center, no on any other part of the metropolitan area.



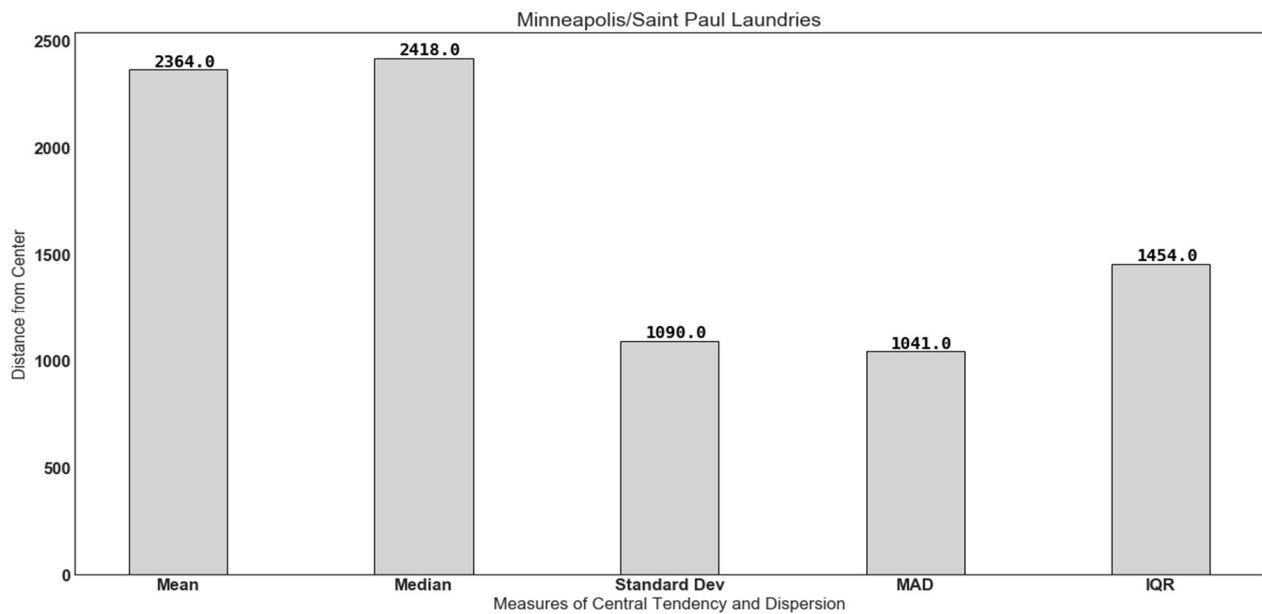
The calculation again used the Halverson formula. And once identified, I calculated the distance from each laundry to the center.

A simple histogram shows the distribution more normally distributed with the most common distances at the center and

shorter and longer distances less frequent. A calculation of measures of central tendency and dispersion show that the average distance of a laundry to the center of laundries is 2,364 meters. The median is similar at 2,418 meters, representing a skew in the distribution. The standard deviation is relatively lower than the Milwaukee standard deviation at 1,090 meters.



The mean absolute deviation, was similar to the standard deviation at 1,041



meters while the inter-quartile range was a little higher at 1,454 meters.

#### 4.3 Comparison between Milwaukee and Minneapolis/Saint Paul Laundries

The median distance from laundries in Milwaukee to the center was 2,196 while the median distance in Minneapolis Saint Paul was similar at 2,418 meters. These two numbers are similar. The standard deviations, at 1,710 in Milwaukee, and 1,090 in Minneapolis. This would suggest that laundries are a lot more dispersed in Milwaukee than in Minneapolis. The mean absolute distance figure also supports this idea. The mean absolute distance figure in Milwaukee as 1,862 and in Minneapolis/Saint Paul was 1,041.

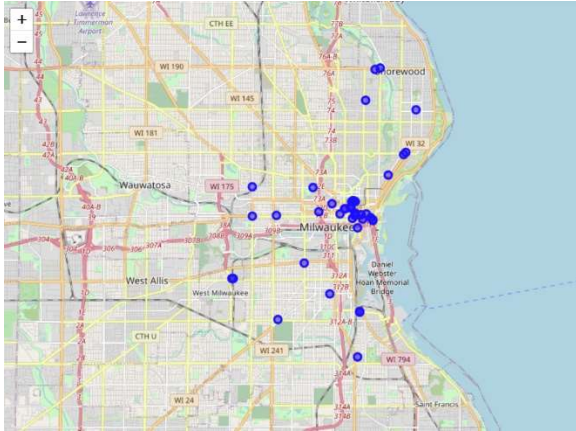
Altogether, the statistics suggest that the central tendency of laundry distances from the laundry center are similar in

Milwaukee and Minneapolis/St. Paul, but the dispersion of laundries is much wider in Milwaukee than in Minneapolis/Saint Paul.

#### 5. Banks

As a first step, I plotted all banks on a Milwaukee map and a Minneapolis map. These maps show something similar and dissimilar. They are similar in that most banks are near the downtown area. They are not in suburban areas. They are dissimilar in that the Milwaukee laundries are spread more north/south, and the Minneapolis/Saint Paul laundries are spread more circularly around the city center. This is similar to the pattern found with laundries, though with Milwaukee banks more spread out on the map and Minneapolis/Saint Paul banks centered pretty tightly on the map.

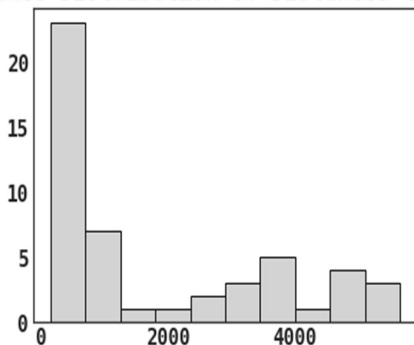




### 5.1 Milwaukee Banks

Foursquare identified 49 banks and financial businesses in Milwaukee. I calculated the geographic center of these banks at latitude 43.0399 and longitude -87.9141. This is close to the city center latitude and longitude of 43.0389 and -87.9065. Again, this indicates that banks are centered right on the city center, not in the suburbs or another part of the metropolitan area. It also means that even with some dispersion, banks are mostly evenly spread from the city center.

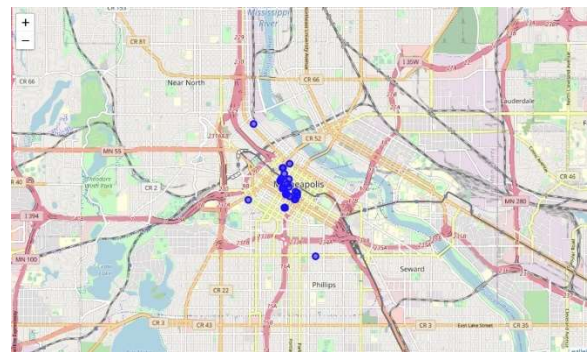
Milwaukee Distribution of Distances to Center



Once the center was identified, I calculated the distance from each

Milwaukee bank to the center. A simple histogram of those distances shows a large number of distances are short with a tail leading toward longer distances.

A calculation of measures of central tendency and dispersion show that the average distance of a bank to the center of banks is 1,887 meters. The median is a much shorter, at 780 meters, representing skew in the distribution. The standard deviation is relatively high in comparison to the mean and median at 1,812 meters. The mean absolute deviation, which was a similar 695 meters while the inter-quartile range was broader at 2,860 meters.

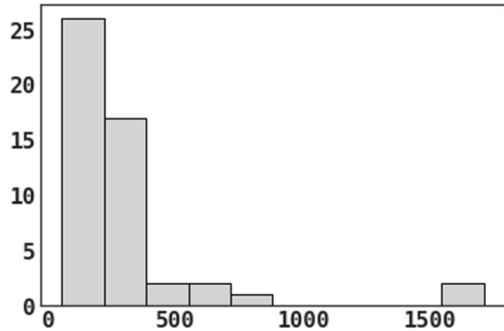


### 5.2 Minneapolis/Saint Paul Banks

Foursquare found 49 banks and financial businesses in Minneapolis/Saint Paul. I calculated the geographic center using the Haversine formula to be latitude 44.9757 and longitude -93.2691. This again is similar to the latitude and longitude of the city center at 44.9778 and -93.2651. This

again confirms that banks center on the city center itself, rather than on some other neighborhood or district.

Distribution of Distances to Center



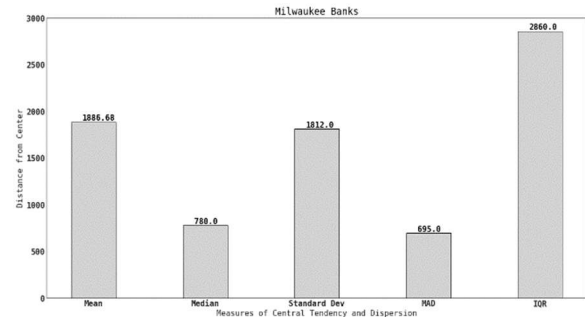
Once the center has identified, I calculated the distance of each bank to the center. A simple histogram of those distances shows that most banks are close to the center with a few banks a much greater distance from the center.

Central tendency and dispersion measures reflect the same thing the histogram shows. The average distance from each bank to the center is 269 meters from the center and the median is even less at 161. The standard deviation is 340 meters and the median absolute deviation is 160. The inter-quartile range is 281.

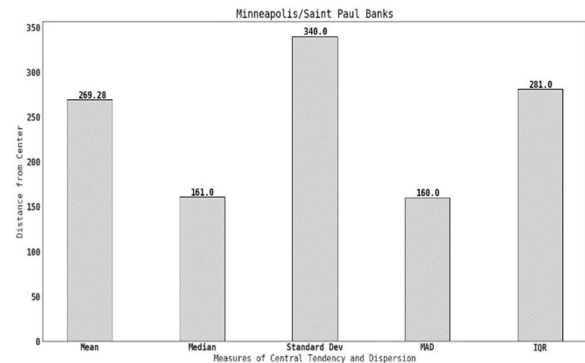
### 5.3 Comparison between Milwaukee and Minneapolis/Saint Paul Banks

Overall, Milwaukee banks are much more spread out compared to Minneapolis banks, as can be seen by the median distances. Milwaukee's median distance of

780 is nearly five times as great as the Minneapolis/Saint Paul median distance of 161.



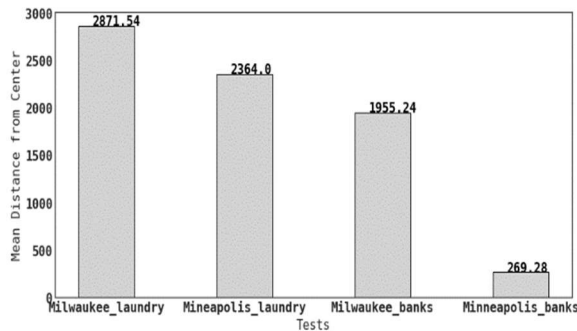
Also, the Milwaukee 695 median absolute deviation is much greater than the Minneapolis/Saint Paul median absolute deviation of 160 (more than four times). Across all measures, the Minneapolis/Saint Paul banks were more tightly clustered than the Milwaukee banks.



## 6. Conclusions

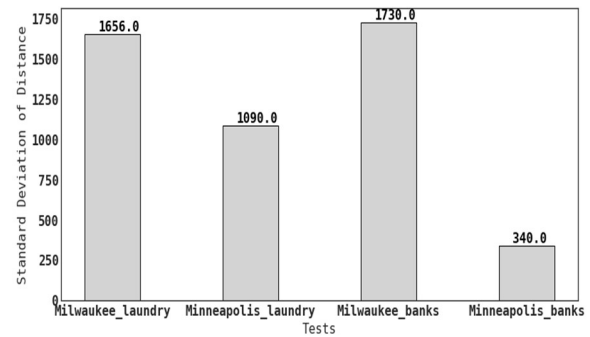
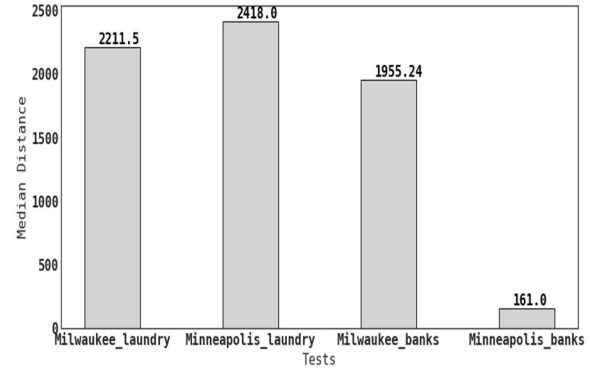
As can be seen by the five summary statistic tables, there are substantial differences both over location and over business type in the average/median distance and the dispersion of those distances.

Compared with the Hotelling thesis that businesses have an incentive to cluster together, the results are mixed. On one hand, both types of businesses in both locations were centered very closely to the city center itself.

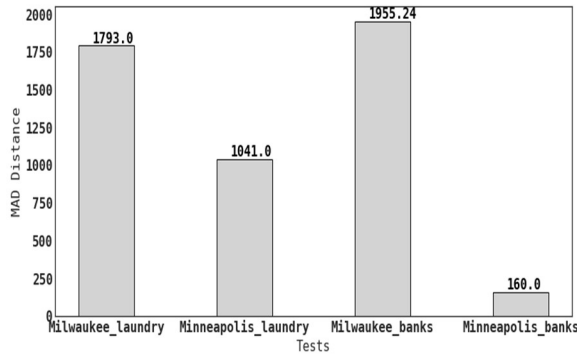


On average, banks and laundries in Milwaukee and Minneapolis/Saint Paul were located around the city center. Banks and laundries were not scattered randomly, they were not centered on a specific district, they were not clustered on a particular neighborhood.

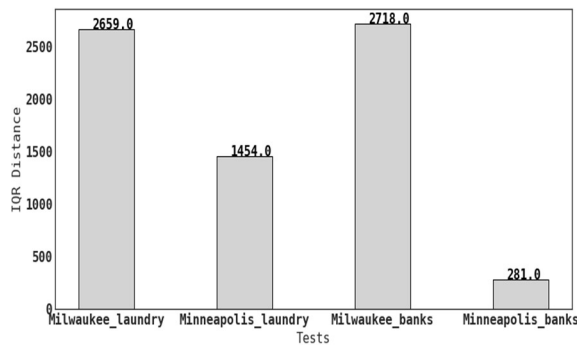
This is all consistent with the Hotelling idea that businesses should locate near the center and near other similar businesses. If I had found that banks or laundries were centered in some other way, maybe by sociographic factors, or according to mass transit, or neighborhood type, that would seem to counter the Hotelling idea. But, instead, both types of businesses in both locations were collected around the city center.



On the other hand, there was marked differences by location and by city. Banks in Minneapolis were much more tightly clustered than any of the other three categories. Also, both Milwaukee laundries and Milwaukee banks were more loosely scattered than either of the Minneapolis/Saint Paul categories. Something about Milwaukee compared to Minneapolis/Saint Paul is related to a greater dispersion of businesses around the city center.



The differences between locations was greater than the differences between business types. While Minneapolis/Saint Paul banks were the most tightly clustered, Minneapolis/Saint Paul laundries were also more tightly clustered.



Again, it appears that the differences by location are greater than the differences by type of business. Going back to the pattern of zip codes on the maps of each state, maybe the shape of the distribution Minneapolis/Saint Paul zip codes radiating from the center shapes how businesses locate in a different way than the distribution of Milwaukee zip codes on a more north/south dimension.

## 7. Future directions

It would be interesting to explore this idea further. Testing mean/median distance and dispersion over a greater number of business types and/or a larger number of cities, might reveal something about what drives some businesses in some cities to cluster more tightly than in others. For the academic, this might provide a fuller explanation of business behavior in determining a location vis-à-vis its competitors.

For a business, it could help inform the strategy of how to best segment a metropolitan area. While it is clear that businesses generally center along the city center, the question is how close should a business approach that center. Is there a middle ground or sweet spot for both carving out a position at the center while also maintaining a differentiation from the competition? Or maybe more data and analysis would suggest that the calculus is different for every city. Once you have decided which city to open a business in, then a calculation can be made on how close to your competitors you should be, and how much of the city you can segment for yourself.