Sushi, Falafel, and Ideology

R. Golan

Introduction

- Sushi made its way to Israel about 30 years ago and quickly gained popularity among the bourgeois society. Being quite expensive, it became a symbol to luxurious food consumed by upper class; and frequently associated with left wing views. Falafel, on the other hand, has always been here. It is considered the Israeli national food and it is filling and very affordable.
- In this capstone project exercise I exploit the data from Foursquare in order to explore the association between the demand to these two very different food options and ideology. My working assumptions include (1) that some degree of spatial equilibrium exists such that businesses are more likely to open where there is higher demand for their services, and (2) that the demand is local by some extent. Aside to the spatial data from Foursquare, I further use the latest election result data from the 2019 Israeli elections.

Data: Description and Exploration

In the following slides I will briefly describe the data and its preparation. These include:

- Official election data (Knesset)
- Geocoding (CBS)
- Sushi and Falafel data (Foursquare)

Official Election Data

- We wish to associate election data with a measure of proximity and access to a sushi place or a falafel stand. We are equipped with official election data, which are specified for each “statistical area” (which is the Israeli equivalent to a census tract, the smallest spatial unit surveyed by the Central Bureau of Statistics, hereafter “CBS”). The complete csv file with the election results can be found here: https://votes21.bechirot.gov.il/ballotresults.
- We add together parties that virtually share ideology, and for normalization we use the share of votes for each group of parties rather than the nominal votes.

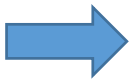
voters	UltOrthodox	Arab	National	Right	Center	Left	statarea	citycode	שם ישוב
62	0.000000	0.612903	0.000000	0.000000	0.241935	0.096774	1.0	967	אבו ג'ווייעד (שבט)
111	0.000000	0.666667	0.027027	0.027027	0.180180	0.045045	2.0	967	אבו ג'ווייעד (שבט)
97	0.000000	0.783505	0.000000	0.000000	0.154639	0.041237	3.0	967	אבו ג'ווייעד (שבט)
191	0.005236	0.450262	0.005236	0.115183	0.151832	0.251309	1.0	472	אבו גוש
287	0.006969	0.536585	0.006969	0.059233	0.139373	0.236934	2.0	472	אבו גוש

- We can see that on average (across pole, not weigted by number of voters), in the average statisttical area 8.2% of votes went to leftwing parties, 25% to the center, 34% to right wing parties, 6.5% to (religious) national parties, and almost 10% to arab parties. Finally, we see that on average 400 people voted in each statistical area.

Geocoding

- To get the central coordinates of each statistical area I turn to the SHP file provided by the CBS that contains the statistical area polygons. as an aproximation to the centroid of each statistical area I use the mean coordinates of the polygon points. That of course is not accurate, but seems reasonable for relatively symetric polygons. (the shp files can be downloaded from www.data.gov.il).
- It turns out, however, that these are provided in a local system (ITM) - see _X and _Y variables. Some non-linear conversion is needed before I can pass them to make a Foursquare api call. To overcome this issue, I wrote functions that converts ITM to the global system and vice-versa, as well as functions that do the same for values stores in DataFrame columns.

YISHUV_STA	STAT08	SEMEL_YISH	OBJECTID	sa_id	Y_	X_	D_
0	0	2-	1	1	727858.153125	215275.551238	1
70001	1	7	2	2	614305.807668	174053.640065	2
100001	1	10	3	3	628677.177597	189331.387874	3
110001	1	11	4	4	612136.421822	173435.339793	4
130001	1	13	5	5	520252.822160	226287.155326	5



lat	lng
32.644746	35.159178
31.619840	34.725228
31.749952	34.885801
31.600251	34.718813
30.772438	35.275106

Sushi and Falafel Data

- This is partially a part of the method, but now I want to characterize every statistical area by some proxy of the access to sushi places and falafel stands. I therefor use the Foursquare API using a query of either “falafel” or “sushi”, and the coordinates of each statistical area. I let the radius and the limit to be high (10km, 100) and then weight each of this venues by the distance to the center of the statistical area. This procedure is common in urban economics, and used in what is called "gravity models". That allow me to have a single and continuous indicator for each of my venues of interest.
- $Gravity_i = \sum_j \frac{1}{d_{ij}}$, where i is and index for the statistical area; j an index for the venue within 10 km from i; and d is the distance between statistical area i and venue j.

	SushiGravity	FalafelGravity	statarea	citycode
1	11.353565	0.000000	1	473
2	11.503648	0.000000	2	473
3	13.516351	0.000000	3	473
7	40.851384	18.785532	1	679
9	0.000000	0.000000	1	819

Method and Results

Step1

- As the purpose is to associate ideology with sushi and falafel, we need first to define a proxy for ideology.
- We can simply use the share of voters to left wing parties, yet this proxy ignores the distribution of votes to other parties in the same pole.
- Another approach can be to “let the data talk”. Namely, use a clustering procedure to identify typologies of votes distributions. Then, examine each of them, see if it makes sense, and which of them fits best to proxy a “lefty” area.

Step2

- The second step would be to check if there is an association between ideology to sushi and falafel. I will use both a graph and a logit regression and see if such association exist. In doing so, and since I look only for association, I ignore other possible explanations (such as different socioeconomic status and so on).

Clustering

K-means clustering using solely the election data.

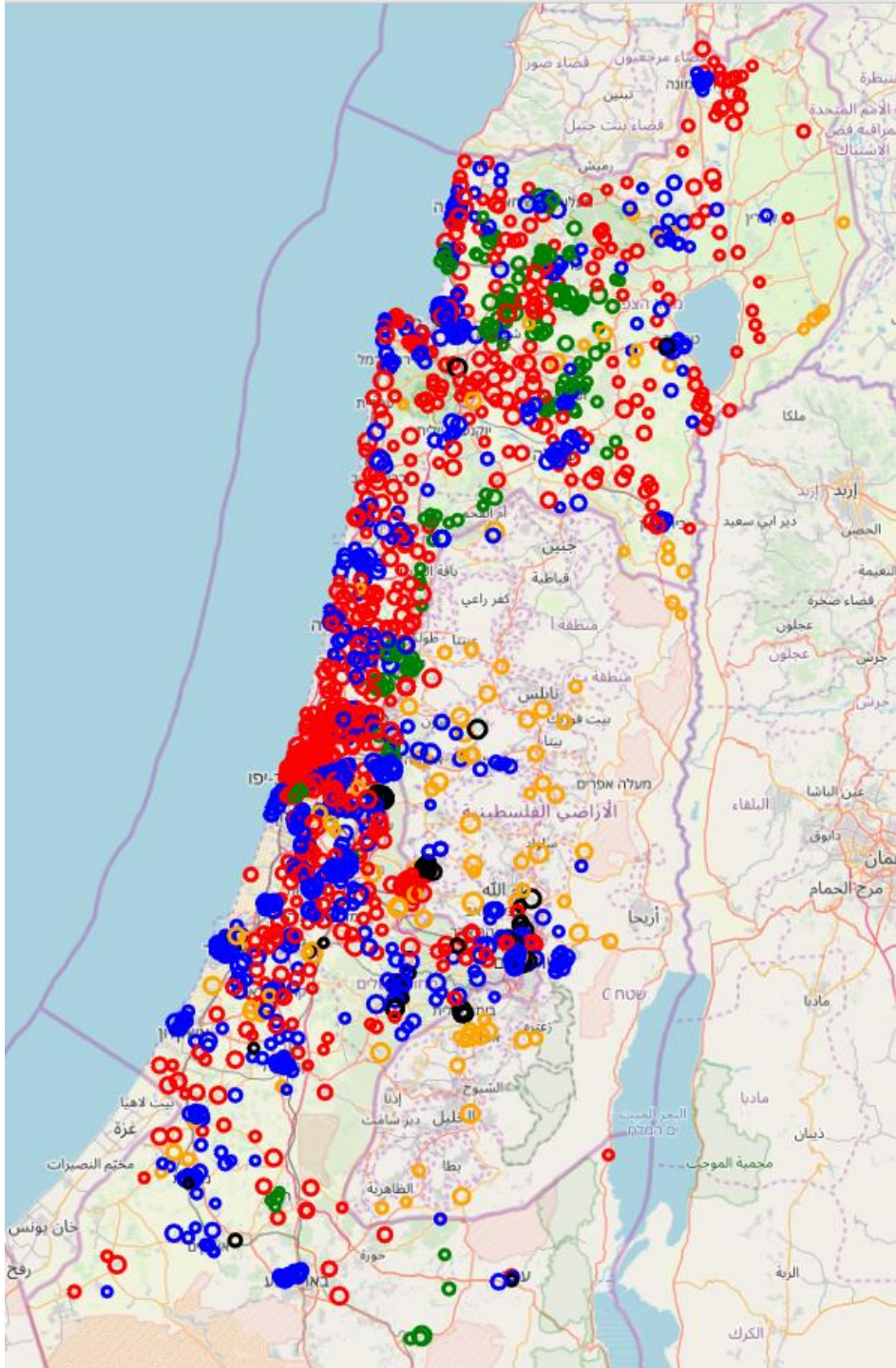
UltOrthodox	Arab	National	Right	Center	Left	Cluster
0.012563	0.007167	0.037760	0.188760	0.502759	0.200208	0
0.100717	0.007590	0.077297	0.535540	0.175755	0.041397	1
0.847325	0.000310	0.059082	0.069204	0.007715	0.001786	2
0.016380	0.786815	0.002478	0.037128	0.040518	0.090549	3
0.042596	0.000314	0.618009	0.202777	0.045298	0.014358	4

Clustering

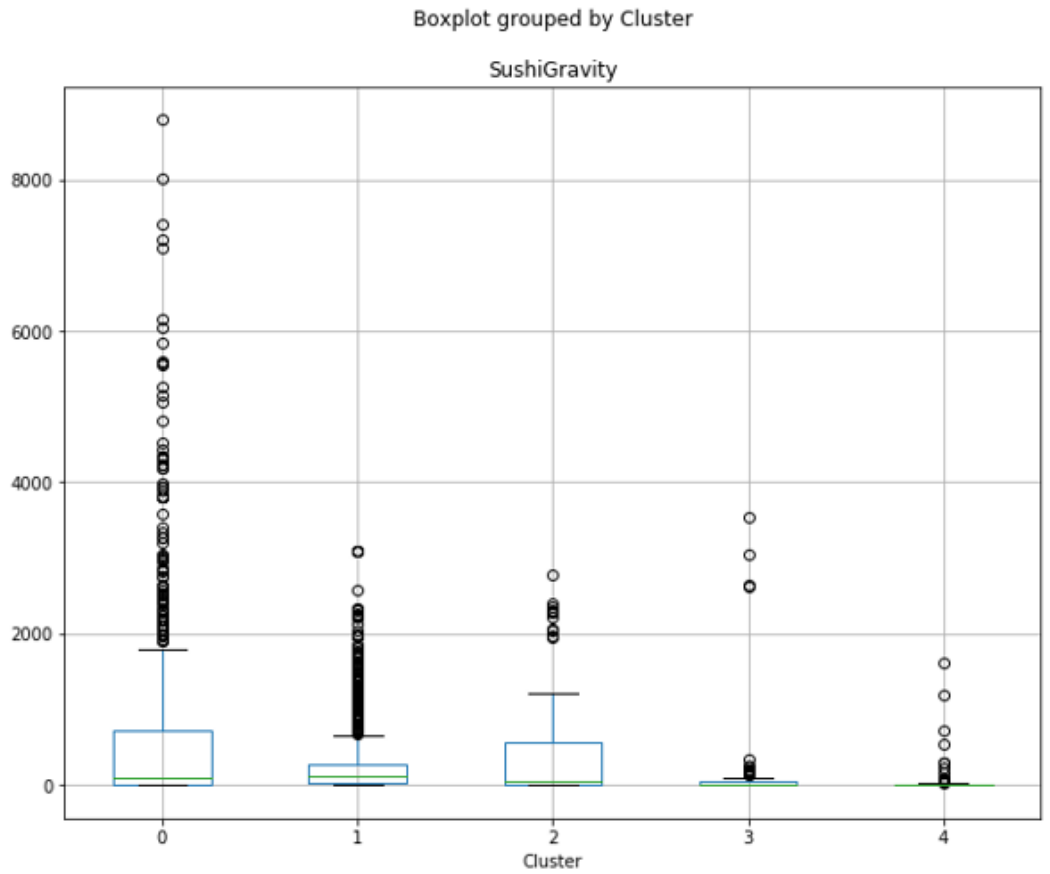
A map using Folium.

The size indicates the number of voters, the color indicates the cluster

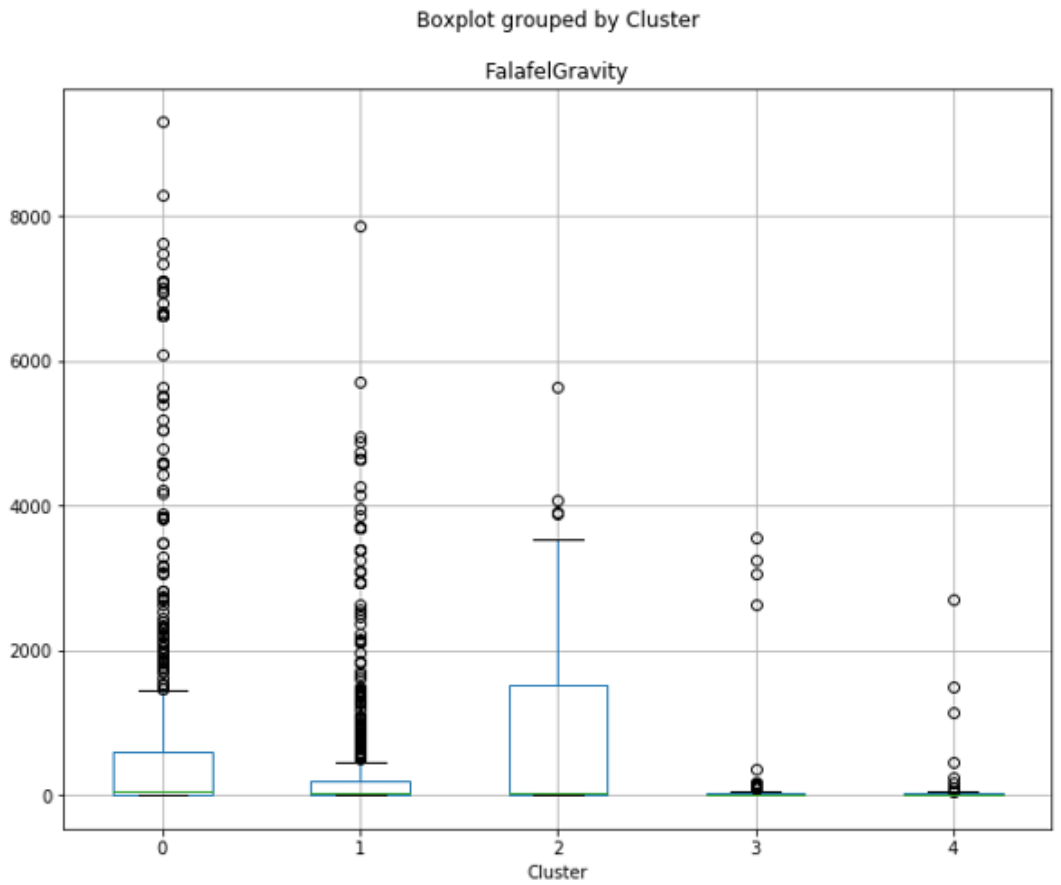
UltOrthodox	Arab	National	Right	Center	Left	Cluster
0.012563	0.007167	0.037760	0.188760	0.502759	0.200208	0
0.100717	0.007590	0.077297	0.535540	0.175755	0.041397	1
0.847325	0.000310	0.059082	0.069204	0.007715	0.001786	2
0.016380	0.786815	0.002478	0.037128	0.040518	0.090549	3
0.042596	0.000314	0.618009	0.202777	0.045298	0.014358	4



Sushi and the election-based cluster



Falafel and the election-based cluster



Sushi, Falafel, and Ideology

```
.Optimization terminated successfully
Current function value: 0.684857
Iterations 4
Results: Logit
================================================================
   Model:              Logit            Pseudo R-squared: -0.017
Dependent Variable: y                AIC:              2409.2170
Date:               2019-06-27 10:58 BIC:              2420.1586
  No. Observations:   1756             Log-Likelihood:   -1202.6
  Df Model:           1                LL-Null:          -1182.5
   Df Residuals:       1754             LLR p-value:      1.0000
   Converged:          1.0000           Scale:            1.0000
                                       No. Iterations:   4.0000
----------------------------------------------------------------
Coef.  Std.E      z     P>|z|   [0.025  0.975]
----------------------------------------------------------------
SushiGravity      0.0006   0.0001  4.2993 0.0000  0.0003  0.0008
FalafelGravity   -0.0003   0.0001 -2.9394 0.0033 -0.0005 -0.0001
================================================================
```

Access to sushi increases the likelihood of a “lefty” statistical area

Access to falafel decreases the likelihood of a “lefty” statistical area

Discussion and Concluding remarks

- In this project I`ve explored ideology viz-a-viz demand for two types of food: one luxurious and the other one basic. Our findings suggest that Sushi in deed is associated with leftwing opinions whereas Falafel is associated with right wing opinions.

- In addition, we have cluster Israeli statistical areas (~neighborhoods) by the elections at the poles using the latest 2019 data. This clustering by itself reveals an interesting agglomerations of ideologies across space in Israel. The ability to intuitively interpret the clustering results increases, in my view, the reliability of this simple procedure.

- The tools that were used in this project include, among others, k-means clustering, folium maps, logit regression and data retrieval from Foursquare using an API. I further developed for this project a function that converts ITM coordinates to the global system.