

Number of Health-Promoting Resources and Life Expectancy at Birth in Each Community of Chicago

Capstone Project for IBM Data Science Professional Certificate on Courser

Nam-kyu Cho

1. Introduction

In a recent study by NYU School of Medicine, Chicago was found to have the broadest gap in life expectancy at birth across neighborhoods among 500 largest cities in the US (NYU Langone News, 2019). There is a 30-year gap in life expectancy between residents living in Streeterville and people living in Englewood, neighborhoods that are only 9 miles apart (Schencker, 2019). There are many possible factors that could contribute to this disparity in life expectancy including differences in education, race/ethnicity, income, and access to health promoting resources (Hunt, Tran, & Whitman, 2015; Singh et al., 2017). These are some of the social determinants of health, which are “conditions in which people are born, grow, live, work, and age”; these are affected by social structures in turn (Marmot et al., 2012).

Among the social determinants, disparities in race/ethnicity appears to have a substantial effect on the gap in life expectancy (Hunt et al., 2015). This, in turn, may be caused by residential segregation (Williams & Jackson, 2005). Chicago, along with other cities with a wide gap in life expectancy, is highly segregated in terms of race and ethnicity (NYU Langone News, 2019). Even after adjusting for socioeconomic status, segregation due to race/ethnicity can have a profound effect on health through differential exposure to physical environment of neighborhoods, poverty, and economic and educational opportunities (White & Borrell, 2011; Williams & Jackson, 2005). Access to health-promoting resources of neighborhoods, including sports facilities, outdoor places, healthcare, and healthy food has been found to be very important for health, with segregation and disparate allocation of resources limiting access to those health-promoting resources and having a negative effect on health (Billaudeau et al., 2011; Coen & Ross, 2006; Inagami, Cohen, Finch, & Asch, 2006; Kawakami, Winkleby, Skog, Szulkin, & Sundquist, 2011).

In this study, to explore this disparity between neighborhoods further, I used the Foursquare API to search for health-promoting venues in each community and clustered the communities using k -means clustering based on the frequency of each category of those venues. Those clusters were compared in terms of life expectancy. The clusters were also mapped on a choropleth map of Chicago communities so that government officials and any stakeholders in the healthcare business of Chicago can identify the areas of need and allocate their resources accordingly.

2. Data

Life expectancy at birth data for each community of Chicago for year 2016 (2017 also exists, but the most recent global data was from 2016) was taken from the Chicago Health Atlas (City Tech Collaborative & Chicago Department of Public Health, n.d.). This dataset was produced using death certificate files provided by the Illinois Department of Public Health (IDPH) with life expectancy calculated using the Chiang methodology. This data was used to compare life expectancy at birth among communities and produce a choropleth map visualizing the disparities. The life expectancy data of various countries including the US for year 2016 was retrieved from WHO Global Health Observatory data repository (WHO, 2018).

The geojson file of Chicago communities used to indicate community boundaries on the maps was taken from the Chicago Data Portal (City of Chicago, 2018). Health resources for each community was searched using the Foursquare API based on the latitude and longitude of each community gathered utilizing the geopy python package (Nominatim and GeoNames) and Google. The list of resources was retrieved on 8/14/19 around 8:00 AM. This data was used to compare the health-resources of each community and cluster them. Foursquare venue categories were taken from the comprehensive list of categories on Foursquare's official website and manually grouped into 4 different types representing access to sports facilities, access to outdoor places, access to healthcare, and access to healthy food.

3. Methodology

3.1 Data cleaning and exploratory analysis

Data was retrieved from the Chicago Health Atlas and cleaned so that the dataset would only have the 2016 data of life expectancy at birth for each community. Using box plot, bar chart, and descriptive statistics tables, the disparity in life expectancy was explored, and life expectancies were compared with those of other nations. The data of life expectancy in 2016 from other countries was cleaned, and all the data points were plotted along with the lowest and the highest life expectancies from Chicago communities data to illustrate their positions compared to the rest of the world.

3.2 Foursquare API and retrieval of health-promoting venues within each community

In order to look at the disparities in health-promoting resources, Foursquare API was used to retrieve venues of certain categories within a set radius around each community's GPS coordinates. I manually went through each category on the Categories page of Foursquare website (<https://developer.foursquare.com/docs/resources/categories>), assigning relevant categories to the

following groups: 1) Access to sports facilities, 2) Access to outdoor places, 3) Access to healthcare, 4) Access to healthy food. For access to sports facilities, all the subcategories under Athletics & Sports category, such as gym/fitness centers, baseball fields, tennis courts, and other similar categories, was used. For access to outdoor places, a selection of categories including park, playground, and bike trail were used. Access to healthcare includes all the subcategories under the Medical Center category except veterinarians. For access to healthy food, categories such as farmers market, grocery store, and organic grocery were selected. When using major categories, all the subcategories were retrieved using Foursquare API since just using the major category would result in number of all the venues from that category being limited by the API's limit for each request.

GPS coordinates of each community were retrieved using the geopy package with Nominatim geocoder. GeoNames was used to fix South Deering, Lincoln Park, Albany Park, Lincoln Square, Logan Square. Coordinates of Jefferson Park, Dunning, North Center, Humboldt Park, and Washington Heights communities were fixed manually using coordinates Google provided for each community. The reason for some coordinates being out of place seems to be because there are community names that are the same as neighborhood names and/or names of other landmarks. Venues within 1250 meters of the GPS coordinates of each community that belong to each of the four groups were searched using Foursquare API. Number of venues belonging to each group was visualized using tables.

3.3 Number of health-promoting resources and life expectancy

Regression plot was used to visualize the relationship between the number of health-promoting resources and life expectancy. The number of resources in each category was also used to draw scatter plots to illustrate the relationship between each category of resources and life expectancy. Correlation between number of health-promoting resources and life expectancy was explored using Pearson's correlation coefficient and testing its significance.

3.4 Clustering and comparison of clusters

The number of venues (health-promoting resources) were normalized using StandardScaler. The elbow method with the sum of squared distances was used to find the optimal number of clusters. Using the normalized number of resources in each category and the number of clusters, the communities were clustered into different groups using k-means clustering. Different clusters were characterized according to the distribution of health-promoting resources. One-way ANOVA was used to see if there is indeed differences in life expectancies of different clusters.

3.5 Map visualization

Folium library was used to create choropleth map of life expectancies. A geojson layer of community boundaries was used to indicate community areas. With O'Hare and Lakeview communities, the names were changed to OHARE and LAKE VIEW to reflect the community names in the geojson file.

Clustering of communities was visualized by using different colored markers to mark communities belonging to different clusters.

4. Analysis and Results

4.1 Exploratory analysis of life expectancy

Before basic exploratory analysis is performed on the dataset, a dataframe was constructed from the data. A dataframe of communities and their life expectancies was constructed by reading the xlsx file from Chicago Health Atlas. The dataframe was cleaned to only include the data for 2016. The community names and their life expectancies are shown in the table below.

Geo_ID	Community	Life Expectancy
1	Rogers Park	75.3
2	West Ridge	79.2
3	Uptown	76.1
4	Lincoln Square	81.7
5	North Center	82.1
6	Lakeview	81.7
7	Lincoln Park	81.3
8	Near North Side	83.3
9	Edison Park	83.4
10	Norwood Park	81.2
11	Jefferson Park	80.1
12	Forest Glen	82.1
13	North Park	82.6
14	Albany Park	79.5
15	Portage Park	80.0
16	Irving Park	79.3

17	Dunning	79.4
18	Montclare	77.0
19	Belmont Cragin	79.7
20	Hermosa	80.8
21	Avondale	80.5
22	Logan Square	80.3
23	Humboldt Park	74.7
24	West Town	80.6
25	Austin	70.7
26	West Garfield Park	69.5
27	East Garfield Park	68.7
28	Near West Side	78.4
29	North Lawndale	70.0
30	South Lawndale	80.7
31	Lower West Side	81.1
32	Loop	82.0
33	Near South Side	79.4
34	Armour Square	80.6

35	Douglas	74.0
36	Oakland	70.1
37	Fuller Park	66.9
38	Grand Boulevard	72.9
39	Kenwood	78.5
40	Washington Park	68.5
41	Hyde Park	81.9
42	Woodlawn	75.2
43	South Shore	71.5
44	Chatham	73.5
45	Avalon Park	75.1
46	South Chicago	72.5
47	Burnside	70.5
48	Calumet Heights	74.5
49	Roseland	72.2
50	Pullman	71.8
51	South Deering	74.7
52	East Side	79.0

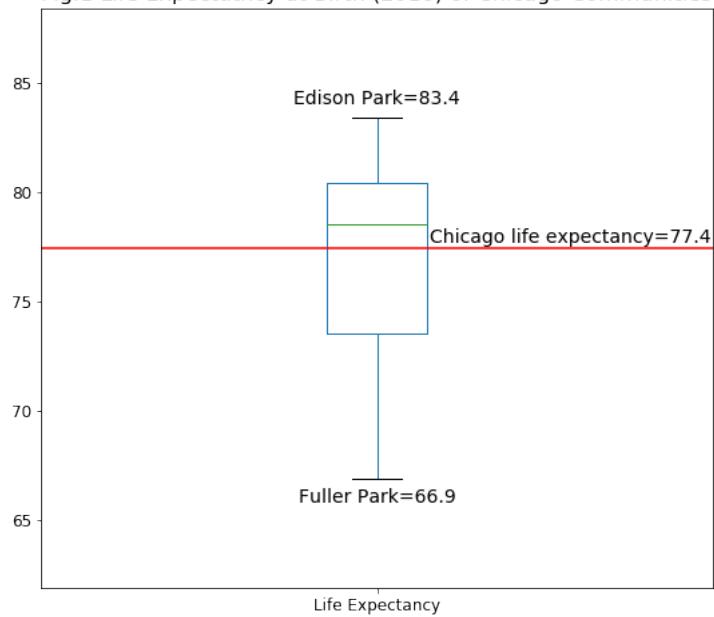
53	West Pullman	72.0
54	Riverdale	71.7
55	Hegewisch	79.1
56	Garfield Ridge	78.6
57	Archer Heights	79.4
58	Brighton Park	79.4
59	McKinley Park	80.4
60	Bridgeport	79.9
61	New City	75.4
62	West Elsdon	79.9
63	Gage Park	81.1
64	Clearing	77.4
65	West Lawn	79.8
66	Chicago Lawn	75.3
67	West Englewood	68.9
68	Englewood	71.5
69	Greater Grand Crossing	70.7
70	Ashburn	76.2

71	Auburn Gresham	72.2
72	Beverly	78.1
73	Washington Heights	74.3
74	Mount Greenwood	78.1
75	Morgan Park	75.2
76	O'Hare	81.8
77	Edgewater	80.3

To explore the characteristics of the data, a box plot (Fig. 1) and a simple descriptive statistics table were used. The box plot shows the disparities in life expectancy at birth across all the communities in Chicago. The red line represents the life expectancy for the whole city, which is 77.4. As shown in the box plot, there is a huge gap between the community with the highest life expectancy, Edison Park with 83.4, and the community with the lowest life expectancy, Fuller Park with 66.9.

The descriptive table shows that the bottom 25% of the communities have life expectancies of below 73.5, which is much lower than Chicago average of 77.4. The top 25% of the communities have life expectancies of above 80.4, which is much higher than Chicago average of 77.4. The difference in life expectancy between Edison Park and Fuller Park was 16.5, which is not as dramatic as the 30-year gap but is still substantial. The median life expectancy is 78.5, which is higher than the life expectancy for the whole city. A little more than half of the communities have life expectancies of more than the city life expectancy. The data looks to be negatively skewed, so the communities in the lower half have life expectancies that are spread farther apart.

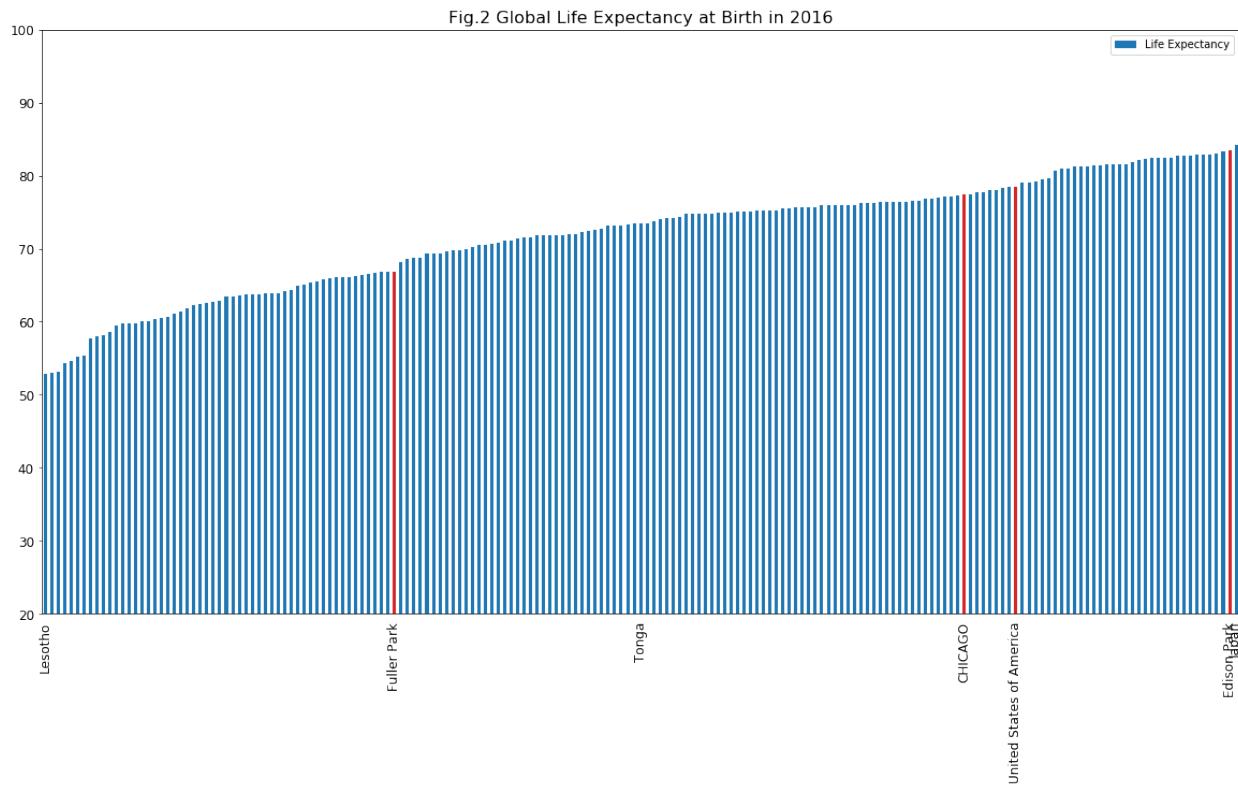
Fig.1 Life Expectancy at Birth (2016) of Chicago Communities



Life Expectancy

count	77.000000
mean	76.922078
std	4.336492
min	66.900000
25%	73.500000
50%	78.500000
75%	80.400000
max	83.400000

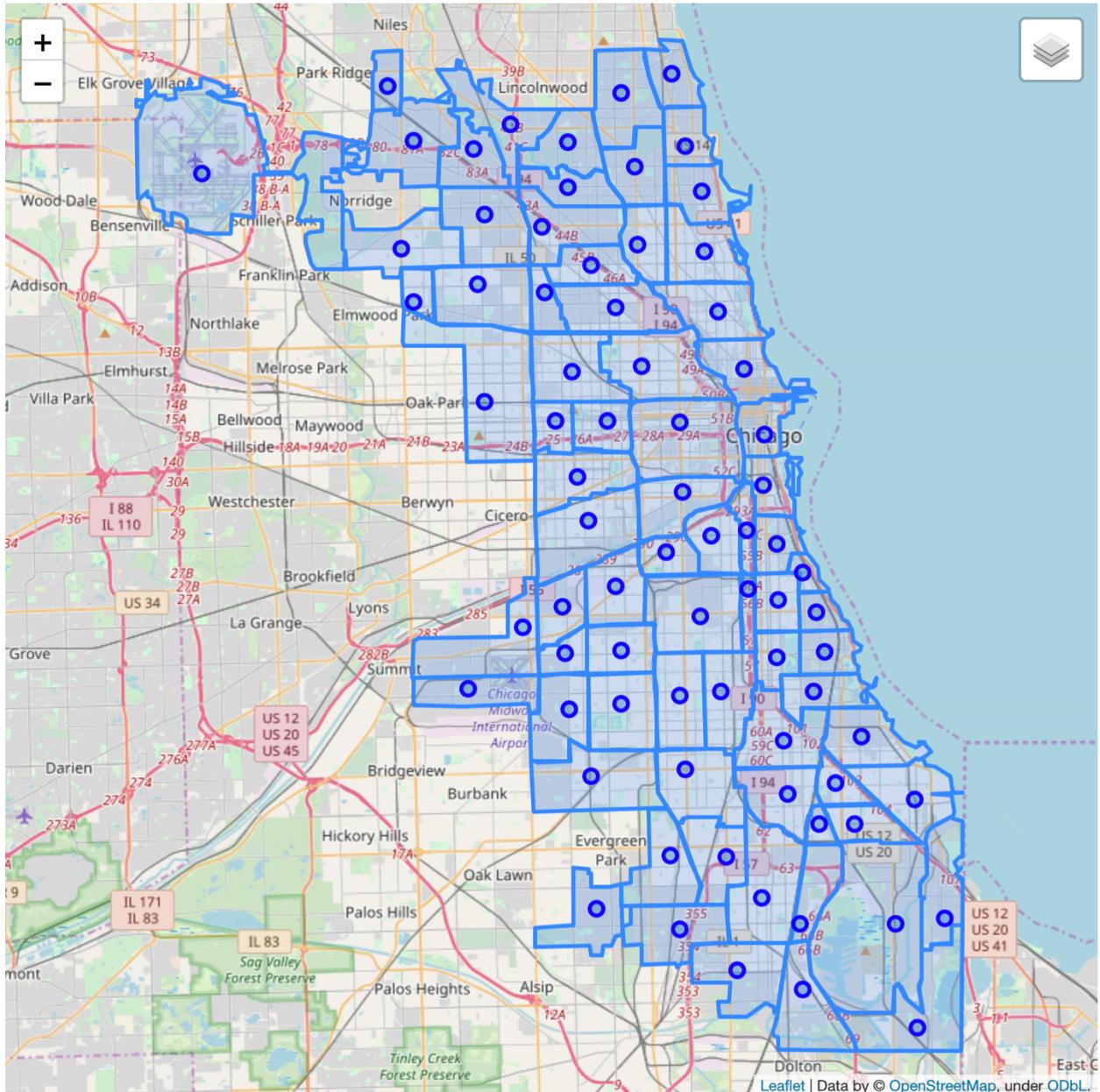
To illustrate how the communities of Chicago compare with other countries in the world in terms of life expectancy at birth, country life expectancy data from WHO Global Health Observatory data repository was used. After cleaning the dataset, the lowest and the highest life expectancies and the average life expectancy from Chicago dataset were combined with the world dataset. Fig.2 shows that while the life expectancy of Edison Park is higher than that of all the other countries in the world except for Japan, and the average life expectancy of Chicago is higher than most countries, the life expectancy of Fuller Park belongs in the bottom half. This chart better illustrates how much difference there between communities of the same city in terms of life expectancy at birth. There are clearly differences in these communities that are affecting life expectancy.



4.2 Distribution of health-promoting resources across communities

To present the possible differences leading to life expectancy, health-promoting resources of each community was gathered. Since each community's GPS coordinates are needed to search for venues in each community, geopy package was used to retrieve the latitude and longitudes of each community as explained in the Methodology section. A geojson file of community boundaries was overlaid to show community areas. The map of Chicago and its communities are shown below. Then, relevant categories of

Foursquare venues were manually put into different groups as explained in the methodology section. A table of full list of categories is shown below.



Access Category	Category Names	ID
0	Sports facilities	Badminton Court 52e81612bc57f1066b7a2b
1	Sports facilities	Baseball Field 4bf58dd8d48988d1e8941735

2	Sports facilities	Basketball Court	4bf58dd8d48988d1e1941735
3	Sports facilities	Bowling Green	52e81612bcbe57f1066b7a2f
4	Sports facilities	Curling Ice	56aa371be4b08b9a8d57351a
5	Sports facilities	Golf Course	4bf58dd8d48988d1e6941735
6	Sports facilities	Golf Driving Range	58daa1558bbb0b01f18ec1b0
7	Sports facilities	Gym / Fitness Center	4bf58dd8d48988d175941735
8	Sports facilities	Hockey Field	4f452cd44b9081a197eba860
9	Sports facilities	Hockey Rink	56aa371be4b08b9a8d57352c
10	Sports facilities	Paintball Field	5032829591d4c4b30a586d5e
11	Sports facilities	Rugby Pitch	52e81612bcbe57f1066b7a2c
12	Sports facilities	Skate Park	4bf58dd8d48988d167941735
13	Sports facilities	Skating Rink	4bf58dd8d48988d168941735
14	Sports facilities	Soccer Field	4cce455aebf7b749d5e191f5
15	Sports facilities	Sports Club	52e81612bcbe57f1066b7a2e
16	Sports facilities	Squash Court	52e81612bcbe57f1066b7a2d
17	Sports facilities	Tennis Court	4e39a956bd410d7aed40cbc3
18	Sports facilities	Volleyball Court	4eb1bf013b7b6f98df247e07
19	Outdoor places	Beach	4bf58dd8d48988d1e2941735

20	Outdoor places	Bike Trail	56aa371be4b08b9a8d57355e
21	Outdoor places	Park	4bf58dd8d48988d163941735
22	Outdoor places	Playground	4bf58dd8d48988d1e7941735
23	Outdoor places	Pool	4bf58dd8d48988d15e941735
24	Outdoor places	Rock Climbing Spot	50328a4b91d4c4b30a586d6b
25	Outdoor places	Ski Area	4bf58dd8d48988d1e9941735
26	Outdoor places	State / Provincial Park	5bae9231bedf3950379f89d0
27	Outdoor places	Trail	4bf58dd8d48988d159941735
28	Healthcare	Acupuncturist	52e81612bcbe57f1066b7a3b
29	Healthcare	Alternative Healer	52e81612bcbe57f1066b7a3c
30	Healthcare	Chiropractor	52e81612bcbe57f1066b7a3a
31	Healthcare	Dentist's Office	4bf58dd8d48988d178941735
32	Healthcare	Doctor's Office	4bf58dd8d48988d177941735
33	Healthcare	Emergency Room	4bf58dd8d48988d194941735
34	Healthcare	Eye Doctor	522e32fae4b09b556e370f19
35	Healthcare	Hospital	4bf58dd8d48988d196941735
36	Healthcare	Maternity Clinic	56aa371be4b08b9a8d5734ff
37	Healthcare	Medical Lab	4f4531b14b9074f6e4fb0103

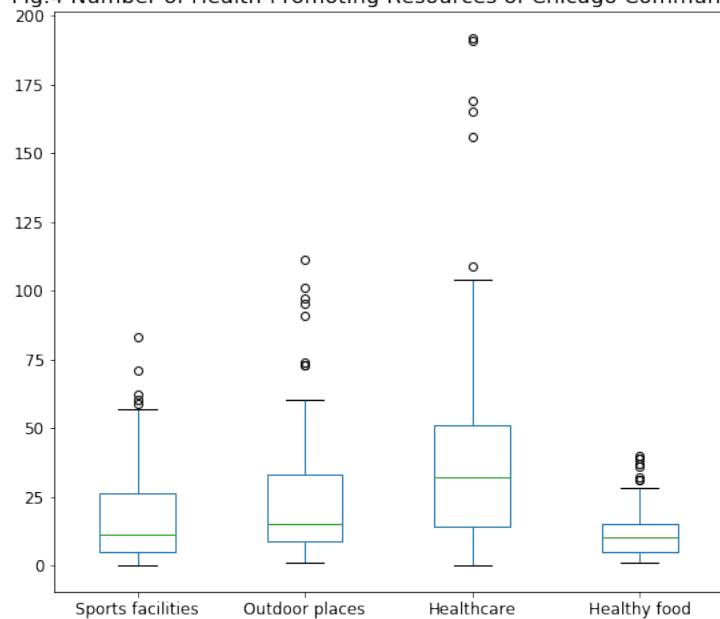
38	Healthcare	Mental Health Office	52e81612bc5c57f1066b7a39
39	Healthcare	Nutritionist	58daa1558bbb0b01f18ec1d0
40	Healthcare	Physical Therapist	5744ccdf4b0c0459246b4af
41	Healthcare	Rehab Center	56aa371be4b08b9a8d57351d
42	Healthcare	Urgent Care Center	56aa371be4b08b9a8d573526
43	Healthy food	Farmers Market	4bf58dd8d48988d1fa941735
44	Healthy food	Fish Market	4bf58dd8d48988d10e951735
45	Healthy food	Grocery Store	4bf58dd8d48988d118951735
46	Healthy food	Health Food Store	50aa9e744b90af0d42d5de0e
47	Healthy food	Organic Grocery	52f2ab2ebcbc57f1066b8b45
48	Healthy food	Supermarket	52f2ab2ebcbc57f1066b8b46
49	Healthy food	Fruit & Vegetable Store	52f2ab2ebcbc57f1066b8b1c

Foursquare API and its search function were utilized to gather venues belonging to each of the four categories. Radius was set to 1250 meters to give it an ample room to gather venues from. Total number of health-promoting resources gathered through Foursquare API was 7732. The healthcare category had the greatest number of venues with 3314, and the healthy food category had the fewest number of venues with 956.

	Healthcare	Healthy food	Outdoor places	Sports facilities	Total
Number of Venues	3314	956	2005	1457	7732

Next, the venue list was grouped by community and access category so that the number of resources for each community were divided into the categories that the resources belong to. The number of health-promoting resources range from 3 (South Deering) to 390 (Loop). That is an extreme difference. Even when the top 10 and the bottom 10 communities are compared, the average number of resources 299.6 and 16, respectively. Across all the categories, the mean was higher than the median, indicating a positive skewness. Top communities in terms of number of resources are even displayed as outliers in the box plot (Fig. 4), further illustrating the disparity. Bottom 25% of the communities have only 5, 9, 14, and 5 or fewer sports facilities, outdoor places, healthcare facilities, and healthy food venues, respectively. On the contrary, top 25% of the communities have 26, 33, 51, and 15 or more sports facilities, outdoor places, healthcare facilities, and healthy food venues, respectively, with top 10 having substantially more. The distribution of healthcare facilities shows the most disparity as the box plot clearly demonstrates. The top 10 communities in terms of total number of resources accounted for almost 40% of the total healthcare facilities. The top 25% of the communities had 4452 resources, which is more than half the number of total health-promoting resources gathered from Foursquare.

Fig.4 Number of Health-Promoting Resources of Chicago Communities



	Community	Sports facilities	Outdoor places	Healthcare	Healthy food	Total
0	Loop	60	101	192	37	390
1	Lakeview	71	111	169	36	387

	Community	Sports facilities	Outdoor places	Healthcare	Healthy food	Total
2	Near North Side	62	91	191	40	384
3	Lincoln Park	55	97	165	28	345
4	Uptown	83	95	76	39	293
5	Edgewater	57	74	101	39	271
6	Near West Side	53	37	156	9	255
7	West Town	51	60	104	32	247
8	North Center	59	54	86	13	212
9	Near South Side	43	73	70	26	212
67	South Chicago	1	9	10	9	29
68	Washington Heights	0	10	10	5	25
69	Pullman	5	8	4	4	21
70	Roseland	2	6	7	5	20
71	West Pullman	3	9	1	4	17
72	O'Hare	7	1	6	2	16
73	Hegewisch	4	4	3	2	13
74	West Englewood	1	7	1	3	12
75	Riverdale	2	1	0	1	4

	Community	Sports facilities	Outdoor places	Healthcare	Healthy food	Total
76	South Deering	1	1	0	1	3

	Sports facilities	Outdoor places	Healthcare	Healthy food	Total
count	77.000000	77.000000	77.000000	77.000000	77.000000
mean	18.922078	26.038961	43.038961	12.415584	100.415584
std	19.369514	26.267571	43.122665	10.136374	92.018927
min	0.000000	1.000000	0.000000	1.000000	3.000000
25%	5.000000	9.000000	14.000000	5.000000	42.000000
50%	11.000000	15.000000	32.000000	10.000000	65.000000
75%	26.000000	33.000000	51.000000	15.000000	121.000000
max	83.000000	111.000000	192.000000	40.000000	390.000000

4.3 Health-promoting resources and life expectancy

I looked at the disparity of health-promoting resources across communities. Then, how does it relate to life expectancy? To compare health-promoting resources and life expectancy across communities, life expectancy was added to the venue list. The table below shows a complete table of number of resources and life expectancy. At a glance, it is clear that the communities with the greatest number of health-promoting resources have life expectancies that are in the high 70s and in the 80s. However, the communities with the fewest number of resources have life expectancies that are in the 60s or in the low 70s with a few exceptions.

	Community	Sports facilities	Outdoor places	Healthcare	Healthy food	Total	Life Expectancy
ID							
32	Loop	60	101	192	37	390	82.0
6	Lakeview	71	111	169	36	387	81.7
8	Near North Side	62	91	191	40	384	83.3
7	Lincoln Park	55	97	165	28	345	81.3
3	Uptown	83	95	76	39	293	76.1
77	Edgewater	57	74	101	39	271	80.3
28	Near West Side	53	37	156	9	255	78.4
24	West Town	51	60	104	32	247	80.6
5	North Center	59	54	86	13	212	82.1
33	Near South Side	43	73	70	26	212	79.4
4	Lincoln Square	39	39	109	11	198	81.7
41	Hyde Park	41	54	85	13	193	81.9
1	Rogers Park	43	73	43	31	190	75.3
22	Logan Square	39	56	68	20	183	80.3
34	Armour Square	31	44	64	19	158	80.6
21	Avondale	40	38	52	26	156	80.5

	Community	Sports facilities	Outdoor places	Healthcare	Healthy food	Total	Life Expectancy
ID							
60	Bridgeport	19	48	46	15	128	79.9
15	Portage Park	26	25	66	10	127	80.0
39	Kenwood	29	53	29	12	123	78.5
2	West Ridge	25	29	36	31	121	79.2
31	Lower West Side	30	33	31	26	120	81.1
16	Irving Park	31	19	49	10	109	79.3
35	Douglas	25	30	42	6	103	74.0
14	Albany Park	22	15	42	21	100	79.5
10	Norwood Park	18	16	64	1	99	81.2
12	Forest Glen	12	29	51	6	98	82.1
20	Hermosa	20	9	46	19	94	80.8
11	Jefferson Park	9	26	55	2	92	80.1
9	Edison Park	24	16	43	8	91	83.4
72	Beverly	12	11	47	15	85	78.1
13	North Park	22	15	43	5	85	82.6
42	Woodlawn	19	20	34	7	80	75.2

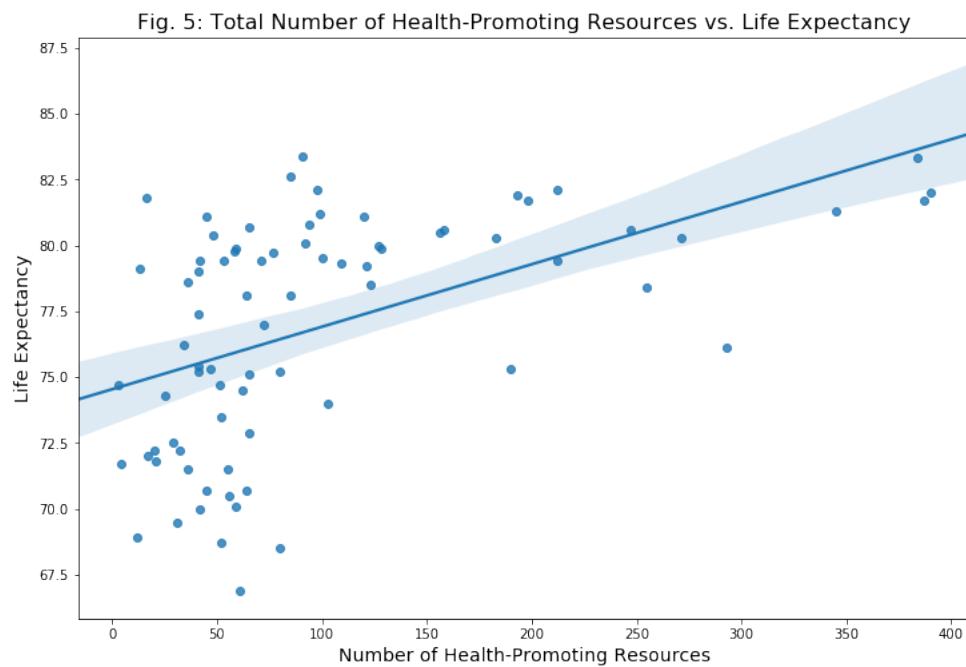
ID	Community	Sports facilities	Outdoor places	Healthcare	Healthy food	Total	Life Expectancy
40	Washington Park	10	14	52	4	80	68.5
19	Belmont Cragin	10	11	39	17	77	79.7
18	Montclare	8	12	37	15	72	77.0
17	Dunning	9	14	42	6	71	79.4
38	Grand Boulevard	7	32	18	8	65	72.9
30	South Lawndale	10	12	31	12	65	80.7
45	Avalon Park	4	6	42	13	65	75.1
25	Austin	14	19	20	11	64	70.7
74	Mount Greenwood	12	10	39	3	64	78.1
48	Calumet Heights	4	5	44	9	62	74.5
37	Fuller Park	18	28	11	4	61	66.9
36	Oakland	10	36	7	6	59	70.1
62	West Elsdon	11	10	31	7	59	79.9
65	West Lawn	11	6	29	12	58	79.8
47	Burnside	5	6	40	5	56	70.5
43	South Shore	13	15	17	10	55	71.5

ID	Community	Sports facilities	Outdoor places	Healthcare	Healthy food	Total	Life Expectancy
58	Brighton Park	4	5	31	13	53	79.4
44	Chatham	4	7	32	9	52	73.5
27	East Garfield Park	13	22	14	3	52	68.7
23	Humboldt Park	5	23	11	12	51	74.7
59	McKinley Park	12	8	11	17	48	80.4
66	Chicago Lawn	7	10	18	12	47	75.3
69	Greater Grand Crossing	4	11	18	12	45	70.7
63	Gage Park	3	8	19	15	45	81.1
57	Archer Heights	7	8	19	8	42	79.4
29	North Lawndale	4	25	9	4	42	70.0
61	New City	2	12	15	12	41	75.4
64	Clearing	6	10	20	5	41	77.4
75	Morgan Park	6	8	21	6	41	75.2
52	East Side	6	15	10	10	41	79.0
56	Garfield Ridge	4	8	23	1	36	78.6
68	Englewood	2	8	18	8	36	71.5

ID	Community	Sports facilities	Outdoor places	Healthcare	Healthy food	Total	Life Expectancy
70	Ashburn	8	12	12	2	34	76.2
71	Auburn Gresham	5	10	7	10	32	72.2
26	West Garfield Park	3	12	10	6	31	69.5
46	South Chicago	1	9	10	9	29	72.5
73	Washington Heights	0	10	10	5	25	74.3
50	Pullman	5	8	4	4	21	71.8
49	Roseland	2	6	7	5	20	72.2
53	West Pullman	3	9	1	4	17	72.0
76	O'Hare	7	1	6	2	16	81.8
55	Hegewisch	4	4	3	2	13	79.1
67	West Englewood	1	7	1	3	12	68.9
54	Riverdale	2	1	0	1	4	71.7
51	South Deering	1	1	0	1	3	74.7

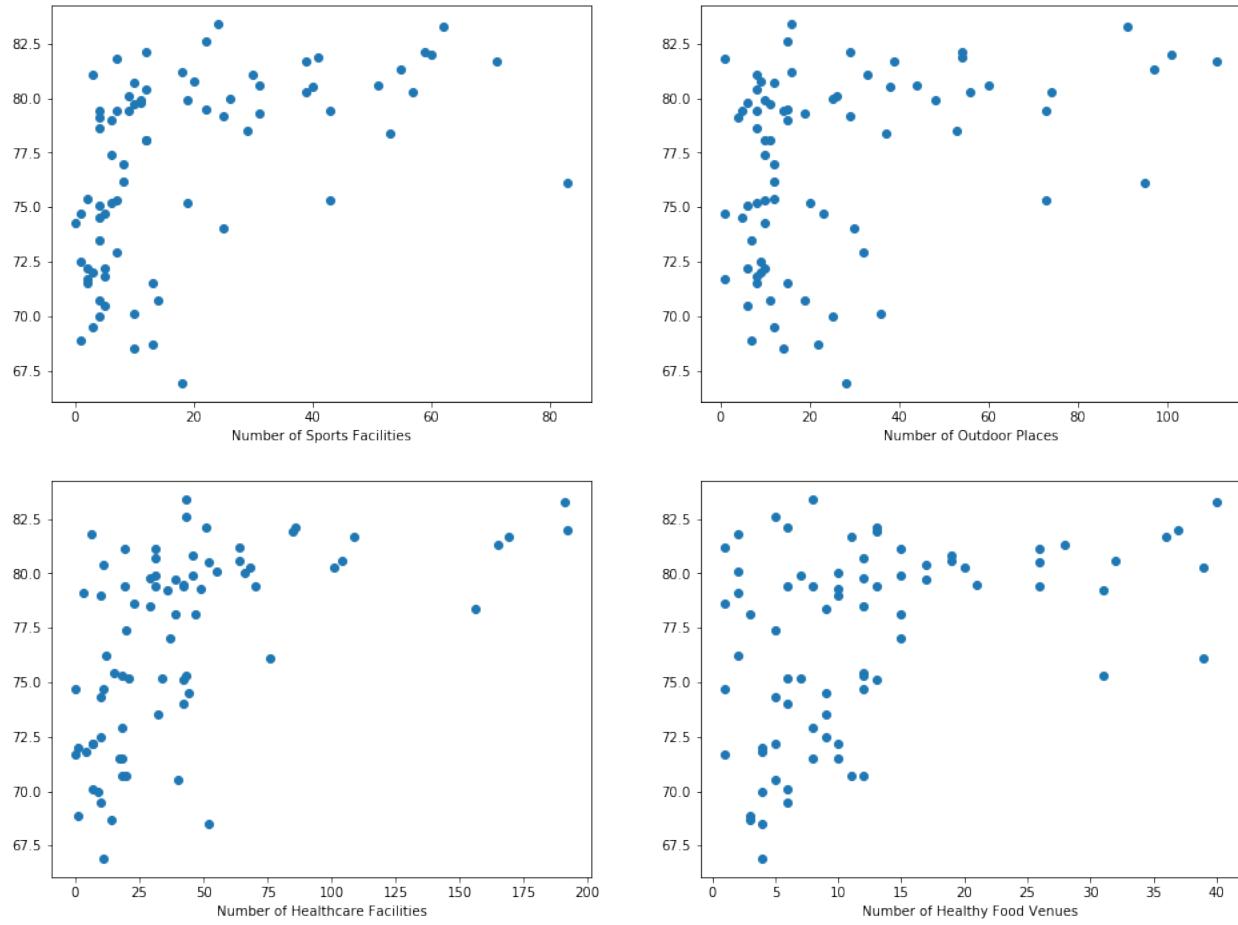
Next, scatter plot (Fig. 5) of total number of health-promoting resources and life expectancy was displayed. Regression line was plotted on the scatter plot to demonstrate the positive relationship between number of resources and life expectancy. However, it looks as if two lines should be fitted as from 0 to about 120 resources, life expectancy increases drastically, but from 100 to 400 resources, life expectancy slowly increases. The Pearson coefficient was 0.503, which shows a moderate correlation between the two variables, but it is statistically significant at $\alpha=0.01$.

Scatter plots (Fig. 6) of each category versus life expectancy were also used to demonstrate the relationship between health-promoting resources and life expectancy. Correlation test was also performed here to test the statistical significance of the correlation. Each category of resources shows positive correlation with life expectancy, but some seem to be more correlated than others. The Pearson coefficients are 0.489, 0.358, 0.534, 0.434 for the 4 categories. Healthcare facilities shows the highest correlation with life expectancy ($r = 0.534$) while outdoor places shows the lowest correlation with life expectancy ($r = 0.358$). They are all statistically significant at $\alpha=0.01$.



The Pearson Correlation Coefficient is 0.5029675906258727 with a P-value of
 $P = 3.13421596734285e-06$

Fig. 6: Number of Health-Promoting Resources vs. Life Expectancy



The Pearson Correlation Coefficient is 0.48881431586861407 with a P-value of
 $P = 6.465453675664565e-06$

The Pearson Correlation Coefficient is 0.35835745521375706 with a P-value of
 $P = 0.0013736271300807054$

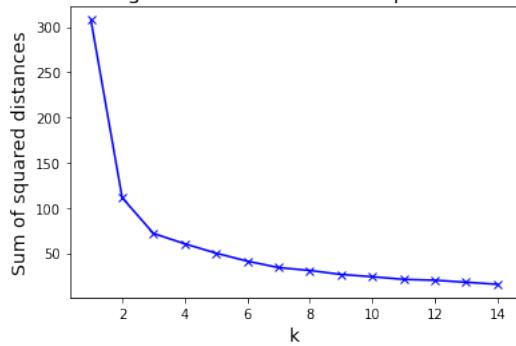
The Pearson Correlation Coefficient is 0.5335061377527421 with a P-value of
 $P = 5.855394655821701e-07$

The Pearson Correlation Coefficient is 0.43359252664773695 with a P-value of
 $P = 8.175033332600771e-05$

4.4 Clustering of communities based on health-promoting resources

To group communities based on health-promoting resources, k-means clustering was used as explained in the methodology section. Using the elbow method (Fig. 7), optimal number of clusters was determined to be 5.

Fig. 7: Elbow Method for Optimal k

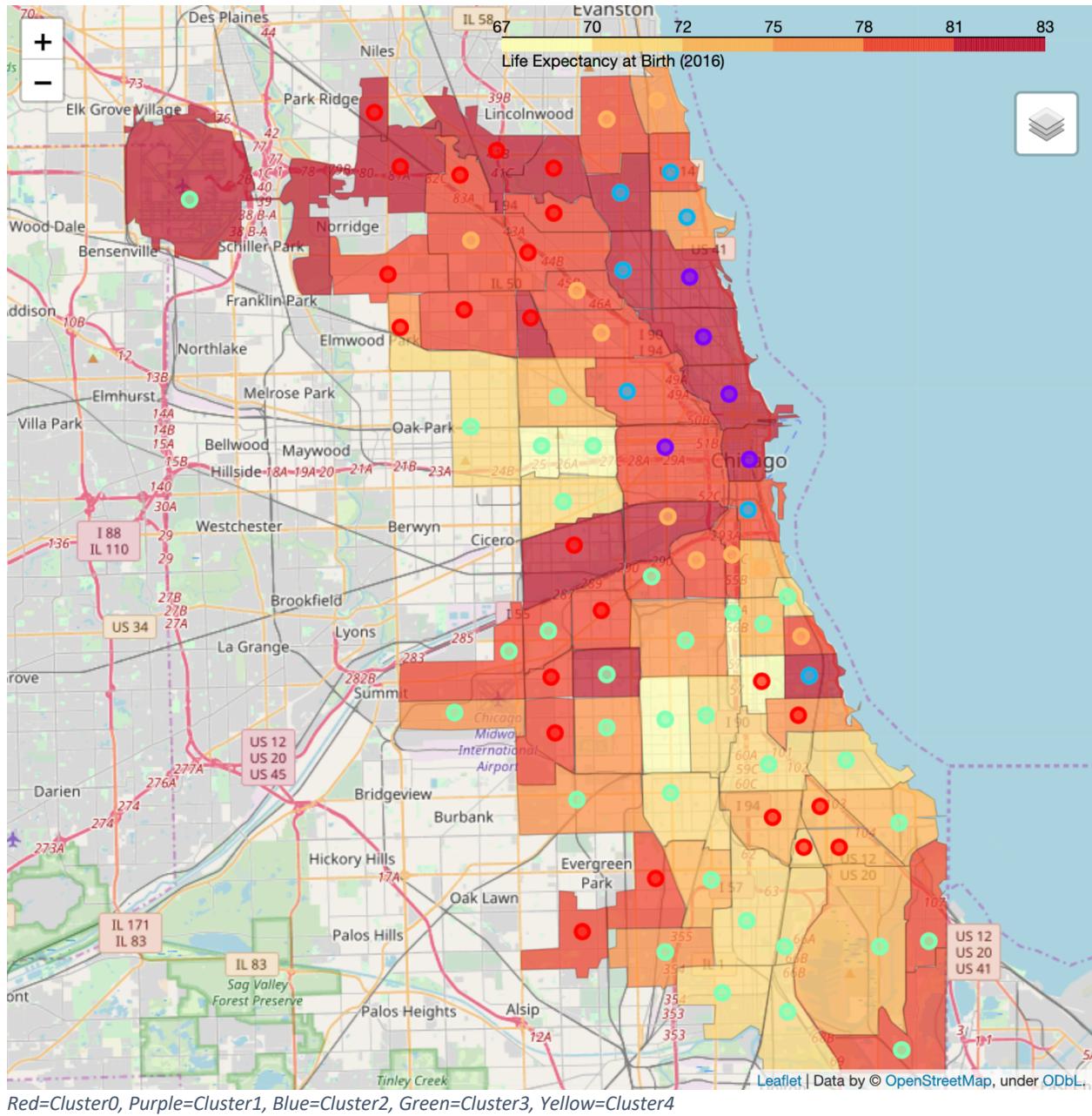


The communities were clustered into 5 different groups. The clusters were straightforward. Cluster 1 had communities with the greatest number of resources. The number of resources in each category are all well above the third quartile and Cluster 2. Cluster 2, with the second highest life expectancy, can also be characterized as having a high number of resources in all the categories. They are all above the third quartile, but much fewer than in cluster 1. With cluster 4, the number of resources in each category are also all a little over the third quartile with the exception of healthcare. The number of resources in communities of cluster 0 are a little over the second quartile for sports facilities and healthcare, but a little below for the other two. Cluster 3, with the lowest life expectancy, can be characterized as having the lowest number of resources across categories. When ANOVA was performed, the result was significant at $\alpha=0.01$, indicating that life expectancy is indeed different across the clusters.

Cluster	Sports facilities	Outdoor places	Healthcare	Healthy food	Total	Life Expectancy
1	60.200000	87.400000	174.600000	30.000000	352.200000	81.340000
2	53.285714	64.142857	90.142857	24.714286	232.285714	80.300000
4	30.700000	42.900000	47.700000	19.600000	140.900000	78.940000
0	12.652174	12.782609	41.869565	9.521739	76.826087	78.191304
3	5.781250	12.375000	11.562500	6.812500	36.531250	73.950000

ANOVA results: $F = 11.113666320045157$, $P = 4.476002426748586e-07$

Finally, a choropleth map with the cluster markers visualize the community clusters with life expectancies. Communities to the west and south have the lowest life expectancies, and most of them belong to cluster 3.



5. Discussions

The communities in Chicago were found to have a 16.5 gap in life expectancy at birth in 2016. It is not as high as the 30-year gap reported in the NYU study, but it is still over one and a half decade. Residents in Fuller Park are expected to live less than 70 years while residents in Edison Park are expected to live well

over 80 years. Such a wide gap became even more apparent when compared with life expectancies of other nations as the life expectancy of Fuller Park is closer to the bottom quarter of the life expectancies of all the countries in the WHO dataset while life expectancies for Chicago and the US are well above the world mean and only Japan has a higher life expectancy of Edison.

In this study, that alarming gap in life expectancy was shown to be significantly associated with access to health-promoting resources of each community. In accordance with previous research, the communities with higher number of health-promoting resources tend to have higher life expectancy at birth. Among the four categories used in the study, access to sports facilities, access to outdoor places, access to healthcare, and access to healthy food, access to healthcare showed the highest disparity, ranging from 0 to 192, which should be a concern for the stakeholders. When the communities were grouped into clusters based on number of health-promoting resources, the clusters also showed a straight-forward trend with the clusters having more resources in each category having higher life expectancies. However, except for Cluster 3 with the lowest number of resources, all the other clusters have average life expectancies of over that of Chicago. The map shows many communities belong to Cluster 3 and are mostly in the south with some in the west. This shows a serious segregation based on the location of communities. Efforts should be concentrated on bringing more health-promoting resources to the communities in Cluster 3 that have the lowest life expectancies. In addition, additional studies should be carried out to figure out how a few communities in Cluster 3 have high life expectancies although they have low number of resources as the other communities in the cluster.

There are some shortcomings of this study that need to be addressed in a future study. I chose communities officially recognized by the city of Chicago. Neighborhoods might be better at representing an actual community that affects a person, but their boundaries often overlap, and there can be different neighborhoods depending on who you ask. However, the boundaries of community areas themselves also present a challenge. Except for a few communities that were annexed afterwards, they were drawn in 1930s by researchers using available data back then. Using the same community areas as a unit of analysis is outdated, but the data was readily available, and it served as a good starting point.

Since I chose to explore venues certain distance away from a single GPS coordinate representing a community, the venues might be underrepresented or overrepresented, and the area surrounding the GPS coordinate may not representative of the community area. In addition, the arbitrary 1250-meter radius should be revised to use an appropriate unit of analysis, which reflects the current situations and how people use the resources around them. Composition of communities should also be taken into

consideration. Some extreme cases include South Deering and Riverdale, with Foursquare venue search returning very few health-promoting resources, which have large percentage of community areas that are parts of marshes, lakes, landfills, and a plant. Another community affected by its lands being occupied by a special facility includes O'Hare as most of its area is occupied by the airport.

In a future study, categories should be further tested and refined. They were constructed based on previously published literature. However, a preliminary analysis should have been performed to choose the categories that can be determinants of life expectancy. In addition to the categories used here, there can be categories of venues a positive effect on mental health, thus prolonging life, such as arts & entertainment places and spiritual center. Types of restaurants can also be considered. There can also be categories of venues that could adversely affect life expectancy that would need to be taken into account such as cigarette shops and liquor stores. The categories included in this study are still quite comprehensive as I used as many as possible from all the Foursquare categories.

Using Foursquare also presented a few challenges. There can also be redundancy in counting the venues of each category in this study. Although users do report duplicate listings, some can still exist, especially since different branches, departments, and sometimes even each hole of a golf course can have their own listings on Foursquare and counted as separate venues. Along with the fact that there are many more venues in areas with denser population, it is possible that the number of resources in the communities with many resources were over-represented. Some venues may not appear on Foursquare because of a number of reasons such as being newly opened or having some changes. There can also be a few that were missed especially in sports facilities category because there can be facilities that can only be categorized as ‘athletics & sports’ and not as any specific sports venue. However, I felt that these discrepancies are not enough to throw the whole picture.

6. Conclusion

There was such a wide disparity in the life expectancies of communities that residents in Edison Park are expected to live more than one and a half decade more than the residents in Fuller Park. Disparity in health-promoting resources seem to have an effect on the gap in the life expectancy. Based on the number of health-promoting resources, including sports facilities, outdoor places, healthcare facilities, and healthy food venues, the communities were clustered into 5 groups. Only one cluster had lower mean life expectancy than life the expectancy of Chicago. That cluster had a substantially low number of health-promoting resources, so stakeholders should improve the infrastructure of those communities to bring in

those resources. In addition, communities with low number of resources but high life expectancy should be studied to see what sets them apart.

7. References

- Billaudeau, N., Oppert, J.-M., Simon, C., Charreire, H., Casey, R., Salze, P., ... Chaix, B. (2011). Investigating disparities in spatial accessibility to and characteristics of sport facilities: Direction, strength, and spatial scale of associations with area income. *Health & Place*, 17(1), 114–121. [https://doi.org/https://doi.org/10.1016/j.healthplace.2010.09.004](https://doi.org/10.1016/j.healthplace.2010.09.004)
- City of Chicago. (2018). Boundaries - Community Areas (current). Retrieved August 9, 2019, from <https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Community-Areas-current-/cauq-8yn6>
- City Tech Collaborative & Chicago Department of Public Health. (n.d.). Life expectancy. Retrieved August 9, 2019, from Chicago Health Atlas website: <https://www.chicagohealthatlas.org/indicators/life-expectancy>
- Coen, S. E., & Ross, N. A. (2006). Exploring the material basis for health: Characteristics of parks in Montreal neighborhoods with contrasting health outcomes. *Health & Place*, 12(4), 361–371. <https://doi.org/https://doi.org/10.1016/j.healthplace.2005.02.001>
- Hunt, B. R., Tran, G., & Whitman, S. (2015). Life Expectancy Varies in Local Communities in Chicago: Racial and Spatial Disparities and Correlates. *Journal of Racial and Ethnic Health Disparities*, 2(4), 425–433. <https://doi.org/10.1007/s40615-015-0089-8>
- Inagami, S., Cohen, D. A., Finch, B. K., & Asch, S. M. (2006). You Are Where You Shop: Grocery Store Locations, Weight, and Neighborhoods. *American Journal of Preventive Medicine*, 31(1), 10–17. <https://doi.org/https://doi.org/10.1016/j.amepre.2006.03.019>
- Kawakami, N., Winkleby, M., Skog, L., Szulkin, R., & Sundquist, K. (2011). Differences in neighborhood accessibility to health-related resources: A nationwide comparison between deprived and affluent neighborhoods in Sweden. *Health & Place*, 17(1), 132–139. <https://doi.org/https://doi.org/10.1016/j.healthplace.2010.09.005>
- Marmot, M., Allen, J., Bell, R., Bloomer, E., Goldblatt, P., & Consortium for the European Review of Social Determinants of Health and the Health Divide. (2012). WHO European review of social determinants of health and the health divide. *Lancet*, 380(9846), 1011–1029. [https://doi.org/10.1016/S0140-6736\(12\)61228-8](https://doi.org/10.1016/S0140-6736(12)61228-8)

NYU Langone News. (2019, June 5). Large Life Expectancy Gaps in U.S. Cities Linked to Racial & Ethnic Segregation by Neighborhood. *NYU Langone News*. Retrieved from <https://nyulangone.org/news/large-life-expectancy-gaps-us-cities-linked-racial-ethnic-segregation-neighborhood>

Schencker, L. (2019, June 6). Chicago's lifespan gap: Streeterville residents live to 90. Englewood residents die at 60. Study finds it's the largest divide in the U.S. *Chicago Tribune*. Retrieved from <https://www.chicagotribune.com/business/ct-biz-chicago-has-largest-life-expectancy-gap-between-neighborhoods-20190605-story.html>

Singh, G., Daus, G., Allender, M., Ramey, C., Martin, E., Perry, C., ... Vedamuthu, I. (2017). Social Determinants of Health in the United States: Addressing Major Health Inequality Trends for the Nation, 1935-2016. *International Journal of MCH and AIDS (IJMA)*, 6(2). <https://doi.org/10.21106/ijma.236>

White, K., & Borrell, L. N. (2011). Racial/ethnic residential segregation: Framing the context of health risk and health disparities. *Health & Place*, 17(2), 438–448. [https://doi.org/https://doi.org/10.1016/j.healthplace.2010.12.002](http://doi.org/https://doi.org/10.1016/j.healthplace.2010.12.002)

WHO. (2018). Life expectancy and Healthy life expectancy - Data by country. Retrieved August 8, 2019, from WHO website: <http://apps.who.int/gho/data/view.main.SDG2016LEXv?lang=en>

Williams, D. R., & Jackson, P. B. (2005). Social Sources Of Racial Disparities In Health. *Health Affairs*, 24(2), 325–334. <https://doi.org/10.1377/hlthaff.24.2.325>