
A Tale of Two Neighborhoods: Toronto & Philadelphia

Data Science Capstone Project
Final Report

A Tale of Two Neighbourhoods: Toronto and Philadelphia

Business Problem/ Objectives

This research was designed to accomplish several goals

Using data science methodology and machine learning techniques like clustering, is it possible to determine whether there is a neighbourhood in Toronto that:

- Offers a similar or better business environment than does Philadelphia to a company like my own?
- Are there any neighbourhoods in Toronto that are similar to my Philadelphia neighbourhood along ‘liveability’ such as: green space, low crime rate, and number of venues for socializing (coffee shops, restaurants, etc.)?
- Are there any neighbourhoods in Toronto that meet both criteria?
- If so, what are they?

For this particular analysis, we will focus on postal/zip code – level data, for those locations that fall within Toronto and Philadelphia city limits.

Target Audience

The primary audiences for this project are:

- The business owner (me);
- Potential investors or other stakeholders; and
- In the event of a decision to move my business to Toronto, Canadian Immigration authorities.

All businesses have start-up costs. Establishing a business in a new location, in a different country, will doubtless carry costs. While bringing such a move to fruition requires a cost estimate, with a list of decision points and criteria, this is beyond the scope of the current undertaking.

Methodology

1. Conduct a review of the relevant literature, using resources available online.
Topics include:
 - Toronto history and current state (geographic, demographic, economic, etc.)
 - Business trends
2. Review Data specifications and availability. Steps include:
 - List data Requirements
 - Collect data – locate Web sites offering Zip and or Postal Code information that can be readily scraped. We will use python’s *beautifulsoup* library to extract postal code lists. Then, we will get the geographical coordinates (latitude and longitude) so we can use them to query the Foursquare API database.¹ A geocoder will allow us to do so. We will then be able to load this information into a pandas dataframe, then using *folium*, we will visualize each city’s neighbourhoods on the map.²
 - Load Foursquare data for all Zip Codes in Philadelphia and all Postal Codes in Toronto. Using the Foursquare API, we will subsequently get the top 100

¹ At this point, we will have set up Foursquare API accounts and gotten Foursquare credentials.

² We will also conduct a ‘sanity check’ to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the cities of Philadelphia PA and Toronto, ON.

venues that are within a radius of 500 meters from the center point of each Zip or Postal Code. We do this by making API calls to Foursquare, passing the geographical coordinates until we are done via a Python loop.

Foursquare then returns venue data to us in a JSON format, and we extract the venue name, category, latitude, and longitude. With these data, we will be able to check to see how many venues were returned for each neighbourhood and to tally up the number of (somewhat)³ unique categories can be curated from all the returned venues.

- Data understanding: Are there any particularities in the data set that must be taken into account during setup (e.g., Zip or Postal codes that fall outside the city limits or are assigned only to P.O. Boxes, and therefore should not be counted)?
 - Clean data – for example, remove from analysis any Zip codes or Postal codes that are actually P.O. boxes.
 - Conduct k-means clustering, using the mean frequency of occurrence of each venue category to create a centroid for each postal code. The k-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as distinct as possible.
 - In the interest of future research, identify and describe limitations of this research, implications, and suggestions for future researchers.
3. Analysis. Perform the analyses listed above, and review information not available through Foursquare about different Toronto neighbourhoods that is not available through Foursquare.
 4. Goal: to identify which neighbourhoods are most likely to meet my twin aims of building a new practice that will offer new services based in Data Science, while permitting me to live someplace like, where I can walk to work.

Data Used

The data for the analyses came primarily from these sources.⁴

- [Foursquare Data](#)
- [Open Street map.org](#)
- [Wikipedia Toronto Web Page](#)
- [Philadelphia Open Access data](#)

Results, Part One: Query and List Zip Codes and Geocoordinates.

Upon querying [OpenDataPhilly](#), among others, we learned that there are 47 Zip codes in the city of Philadelphia, Pennsylvania, USA. We were able to gather the latitude and longitude for each from [zipdatamaps.com](#) and to merge the two in one pandas dataframe.

³ These data are crowd sourced, and the categories are – it seems – far from orthogonal. For example, one category is “food,” which could mean any establishment that sells food. How one distinguishes “food” from “grocery store” is a mystery. See: [Using Foursquare place data for estimating building block use](#),

⁴ This is the short list.

0	Zip Code	Latitude	Longitude
1	19102	39.952962	-75.16558
2	19103	39.952162	-75.17406
3	19104	39.961612	-75.19957
4	19106	39.951062	-75.14589
5	19107	39.952112	-75.15853
6	19111	40.057661	-75.08018
7	19112	39.895677	-75.19044
8	19114	40.064257	-75.00155
9	19115	40.09261	-75.04118
10	19116	40.117413	-75.0154
11	19118	40.07236	-75.20772
12	19119	40.053511	-75.18858
13	19120	40.033944	-75.12118
14	19121	39.981062	-75.1745
15	19122	39.977662	-75.14336
16	19123	39.964012	-75.14764
17	19124	40.017362	-75.08769
18	19125	39.978162	-75.12565
19	19126	40.055411	-75.13793
20	19127	40.026626	-75.22311
21	19128	40.038944	-75.22104
22	19129	40.011562	-75.1839
23	19130	39.968262	-75.17222
24	19131	39.981112	-75.22486
25	19132	39.995412	-75.16977
26	19133	39.992862	-75.14054
27	19134	39.991712	-75.11116
28	19135	40.023611	-75.04966
29	19136	40.041111	-75.02644
30	19137	40.000262	-75.07404
31	19138	40.055861	-75.15654
32	19139	39.961812	-75.23003
33	19140	40.012212	-75.14503
34	19141	40.035778	-75.1447
35	19142	39.922612	-75.23453
36	19143	39.944162	-75.22718
37	19144	40.034111	-75.17203
38	19145	39.922262	-75.18259
39	19146	39.938512	-75.18067
40	19147	39.936562	-75.15409
41	19148	39.919812	-75.15803
42	19149	40.037711	-75.06658
43	19150	40.07226	-75.17106
44	19151	39.975929	-75.25256
45	19152	40.059611	-75.04837
46	19153	39.898985	-75.23221
47	19154	40.09146	-74.977

These are the 47 Zip codes in Philadelphia County, Pennsylvania, USA. Note that Philadelphia is a city that is also a city, which is somewhat unusual. According to the US Census, Philadelphia – the city – has a population of 1,584,138.⁵ The Philadelphia Metropolitan Area – the city and its suburbs – has a population of 6,070,500. This makes it the seventh largest metropolitan area in the country.

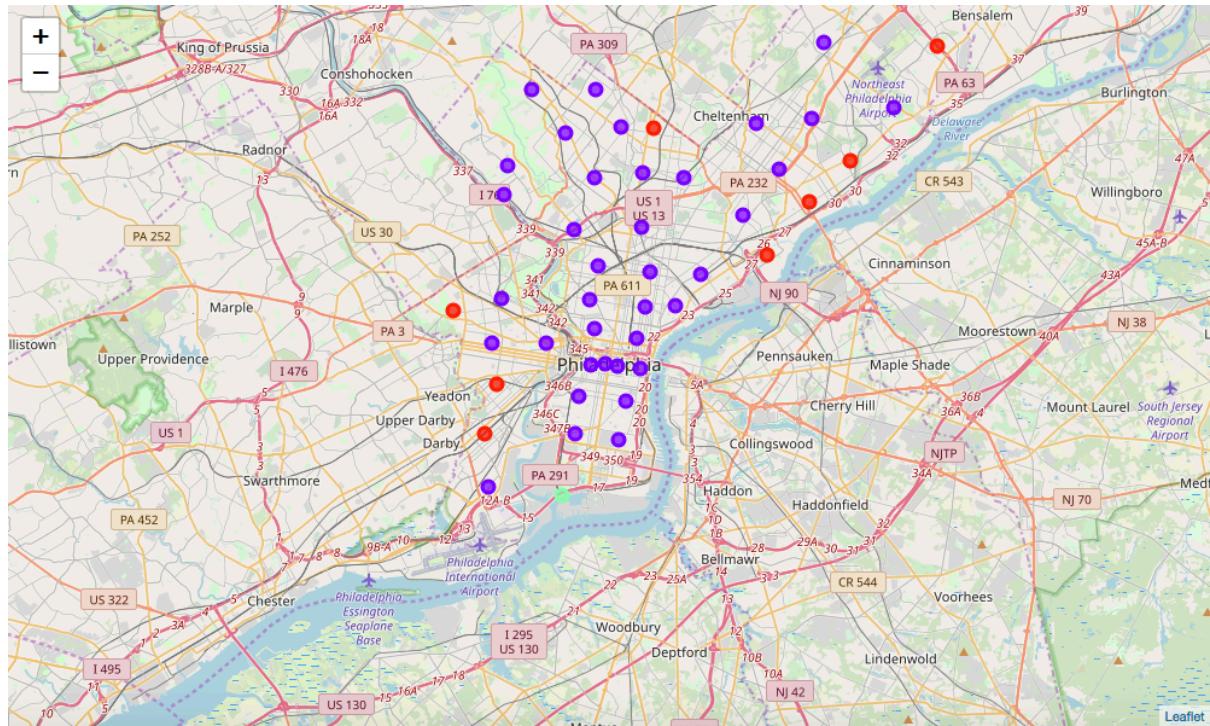
Toronto – the city – had a population of 2,731,571 in 2016. At that time, the Toronto metropolitan area's population was 5,928,040 – although it is estimated to have grown substantially since that time.⁶

⁵ Source: US Census Bureau. Population estimate is as of 2018.

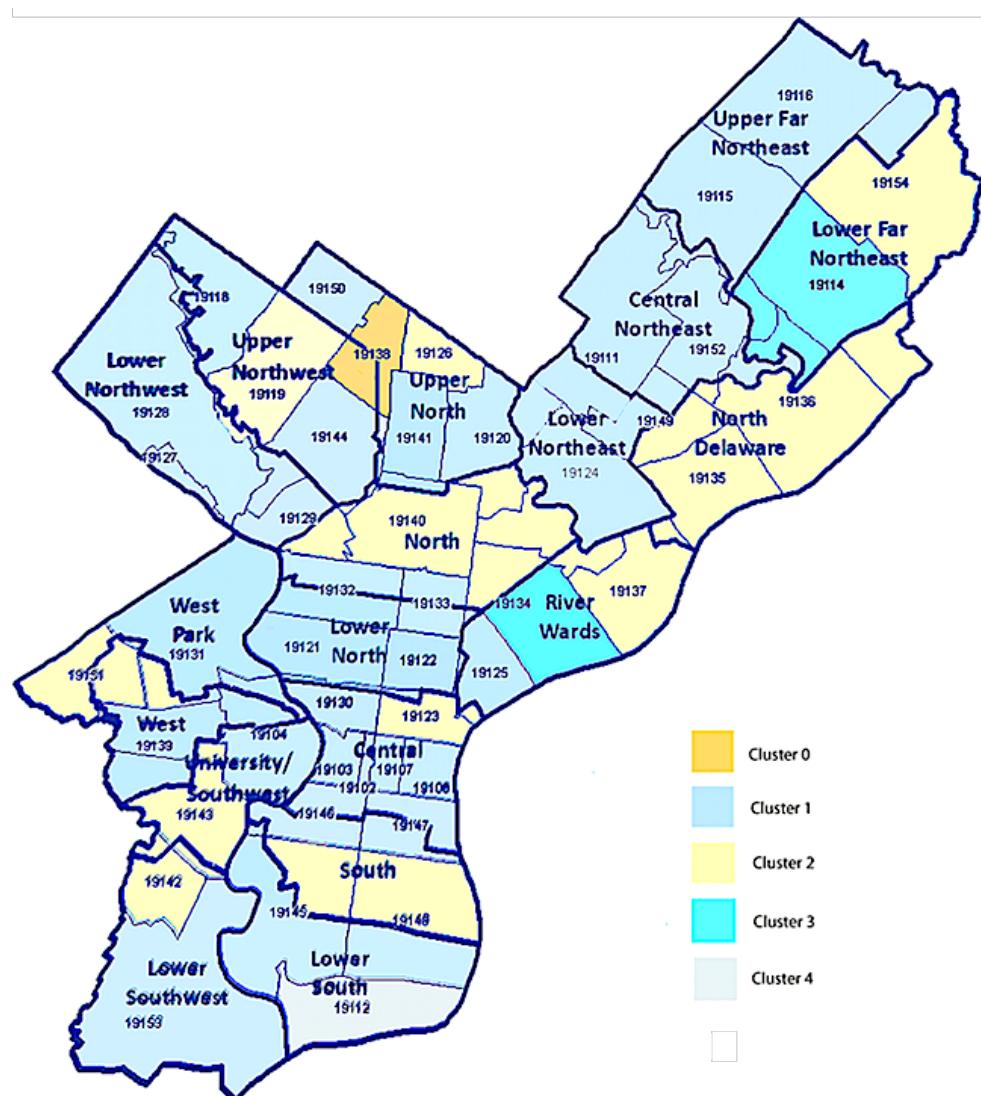
⁶ Source: Statistics Canada.

Results, Part Two: Query Foursquare database, then list venues by Zip code, for all Philadelphia Zip codes, taking a closer look at 19118, Chestnut Hill.

This map depicts the output of the clustering.



This map shows the locations of the five clusters.



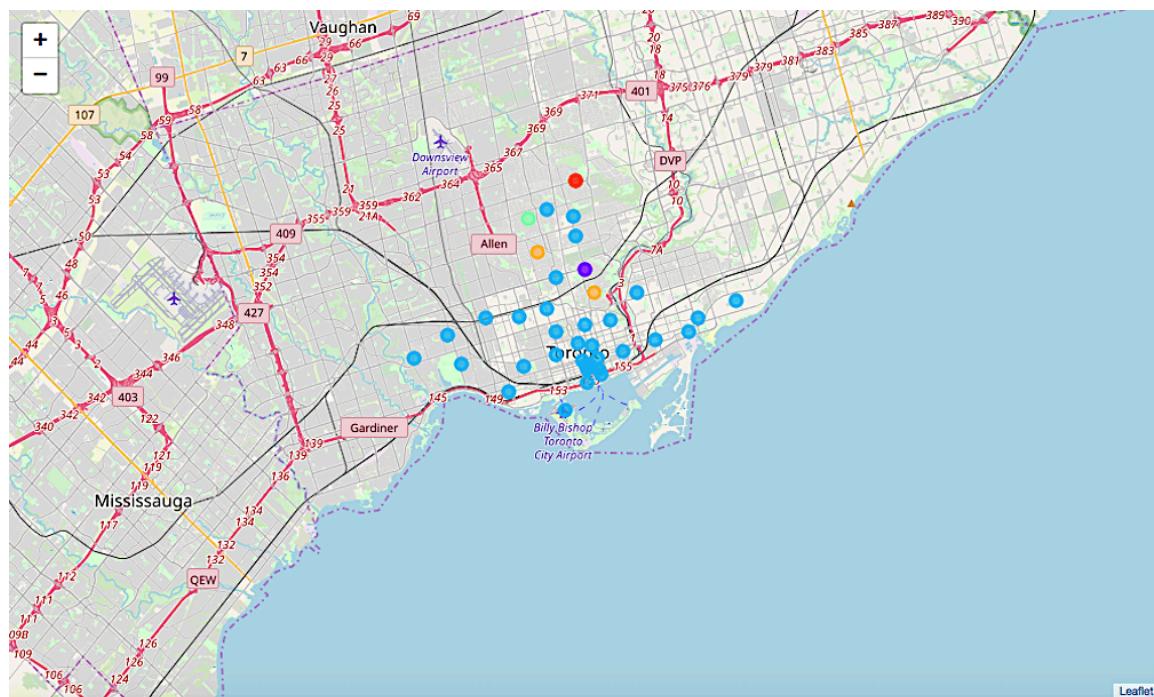
Results, Part Three: Take a closer look at Zip code 19118 – Chestnut Hill.

It appears that the residents of Zip code 19118 like sweets (bakeries, ice cream parlours abound). There are many parks in Chestnut Hill, and the two light rail (commuter) trains suggest that a good number may commute downtown.

Venue Number	Venue Name	Venue Category
5	Campbell's Place	American Restaurant
9	Chestnut Hill Grill	American Restaurant
25	Mica	American - Fusion Restaurant
16	Ten Thousand Villages - Chestnut Hill	Arts & Crafts Store
6	Bredenbeck's Bakery	Bakery
7	Cake	Bakery
24	TD Bank	Bank
29	Wells Fargo	Bank
34	Artisans on the Avenue	Boutique
2	Iron Hill Brewery & Restaurant	Brewery
11	Chestnut Hill Brewing Company	Brewery
31	Roller's Express-O	Café
45	Calypso	Caribbean Restaurant
15	Chestnut Hill Cheese Shop	Cheese Shop
27	King's Garden	Chinese Restaurant
28	Greene Street Consignment	Clothing Store
35	J.McLaughlin	Clothing Store
12	Starbucks	Coffee Shop
40	Poppy's Cafe	Coffee Shop
33	Weavers Way Next Door	Organic Health and Beauty
38	Drake's Gourmet Foods & Catering	Deli / Bodega
13	The Chestnut Hill Farmers Market	Farmers Market
23	Robertson's Flowers & Events	Flower Shop
10	Paris Bistro & Jazz Cafe	French Restaurant
1	Weavers Way Co-Op Chestnut Hill	Organic Grocery Store
3	The Fresh Market	Organic Grocery Store
8	Bredenbeck's Ice Cream	Ice Cream Shop
21	Chill On The Hill	Ice Cream Shop
37	The Knit Wit	Knitting Store
20	Clover Market	Market
26	Jos. A. Bank Clothiers Inc.	Business Clothing Store
4	El Poquito	Mexican Restaurant
41	The Paperia	Paper / Office Supplies Store
0	Pastorius Park	Park
32	The Bone Appetite	Pet Store
22	Fiesta Pizza III	Pizza Place
19	Hideaway Music	Used Record Shop
14	Flying Fish	Seafood Restaurant
18	A Taste of Philly	Snack Place
30	3000BC WellMed Spa	Spa
17	Osaka	Sushi Restaurant
44	Sue's Custom Dressmaking & Tailor Shop	Tailor Shop
36	Stagecrafters	Theater Compaany
43	Oxford Circus Toys	Toy / Game Store
39	SEPTA Chestnut Hill West Station	Local Train Station

Results, Part Four: Take a closer look at Toronto and its Neighbourhoods.

Generate a cluster map for Toronto.



Identify the top venues by neighbourhood cluster. There are 37 boroughs and 240 unique categories: this is a very large dataframe.

Content analysis: how many unique categories can found in the list of venues?

```
In [169]: print('There are {} unique categories.'.format(len(venues_df['VenueCategory'].unique())))
There are 240 unique categories.
```

Below, see a sample of the output from the Toronto clustering analysis.

	PostalCode	Borough	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	M4E	East Toronto	The Beaches	Health Food Store	Other Great Outdoors	Trail	Neighborhood	Pub	Ethiopian Restaurant	Electronics Store	Event Space	Falafel Restaurant
1	M4K	East Toronto	The Danforth West, Riverdale	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop	Furniture / Home Store	Restaurant	Bakery	Pub	Juice Bar
2	M4L	East Toronto	The Beaches West, India Bazaar	Park	Fast Food Restaurant	Ice Cream Shop	Brewery	Liquor Store	Fish & Chips Shop	Burger Joint	Burrito Place	Steakhouse
3	M4M	East Toronto	Studio District	Café	Coffee Shop	Gastropub	Italian Restaurant	Bakery	American Restaurant	Yoga Studio	Gym / Fitness Center	Brewery
4	M4N	Central Toronto	Lawrence Park	Park	Swim School	Bus Line	Yoga Studio	Dog Run	Filipino Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant
5	M4P	Central Toronto	Davisville North	Playground	Hotel	Food & Drink Shop	Clothing Store	Park	Sandwich Place	Grocery Store	Gym	Breakfast Spot
6	M4R	Central Toronto	North Toronto West	Coffee Shop	Yoga Studio	Bagel Shop	Park	Clothing Store	Chinese Restaurant	Rental Car Location	Dessert Shop	Diner
7	M4S	Central Toronto	Davisville	Pizza Place	Dessert Shop	Sandwich Place	Café	Thai Restaurant	Restaurant	Italian Restaurant	Sushi Restaurant	Coffee Shop
8	M4T	Central Toronto	Moore Park, Summerhill East	Tennis Court	Playground	Comic Shop	Discount Store	Filipino Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space

Discussion

The primary purpose of this exercise was to determine whether we were able to use what we learned during the course of this Specialization, independently and without any Lab to provide explicit instructions. In that, this project was successful. I was able to run code that produced a coherent result. I also got to learn more about my own Philadelphia neighbourhood, since that was the topic upon which I chose to focus.

Still, I set out to answer another question, which was "If I were to move to Toronto, is there a neighbourhood that is similar to mine, in a location that would also be within walking distance to an area that could support a growing Consulting/Data Science business? The characteristics of such a location would include mixed residential and commercial use, in a neighbourhood that is roughly as old as my own (founded before 1800), with trees and good places to go (again, on foot) as well as large business's headquarters and growing tech investment.

In both cases, I used Foursquare data to assess the number and type of venues in both cities. However, Foursquare was developed to help people pick somewhere to go, based on their wishes and their location. It was not designed to help people make momentous life decisions. That said, venues matter - and sometimes they can tell you something about a neighbourhood that you might like or really not like. I found this to be the case when I first collected a list of all the venues in my neighbourhood (Philadelphia's Chestnut Hill, Zip code 19118) and the Foursquare Toronto data, parsed by Zip code, latitude and longitude. Please see the spreadsheets I downloaded from my Jupyter notebook, below.

Implications and Conclusions

Relocation analysis is serious business, and this data collection/analysis process is a good beginning. Going forward, I plan to use it as a jumping off point for looking at neighbourhoods, using units of measurement (e.g., the Census block group that are more stable and are linked with larger data sets like the Economic Census, as well as differences in governance, etc. - which are a function of differences between Canada and the US. One thing I know and like very much is Canada's approach to immigration (in other words, it is good for society, the economy, the wellbeing of all and should be encouraged).

Still, I did find a partial answer to my question of where I would want to live/work if I moved to Toronto. "Old Toronto" looks very attractive to me for so many reasons - not the least of which is the presence of the University nearby. Thriving educational institutions are essential to a good economy - especially if the type of work one does is cognitive in nature, as is the case for me. This is a pretty large area and includes a broad variety of neighbourhoods.

One of the analyses I conducted was to look at the mix of venues in my own neighbourhood, then sort the Toronto data to see which among the neighbourhoods covered in our class's work was most similar to my own. One thing I noticed: Chestnut Hill likes food, and parks! There are several ice cream parlours and bakeries a farmer's market, and two of the three grocery stores are organic. There are numerous parks, one of which is among the largest in the US, and there are two light commuter rail lines. Among the Toronto Postal Codes covered by our analysis of Foursquare venues, the Toronto neighbourhoods called Rosedale and Moore Park seem to have the qualities I

would seek. Yes, more research is required, but Data Science has given me more and better tools. This is just the beginning.

Other Information

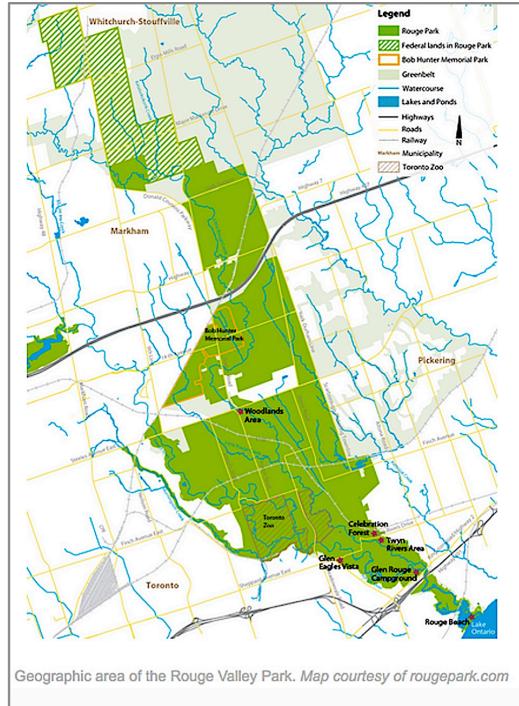
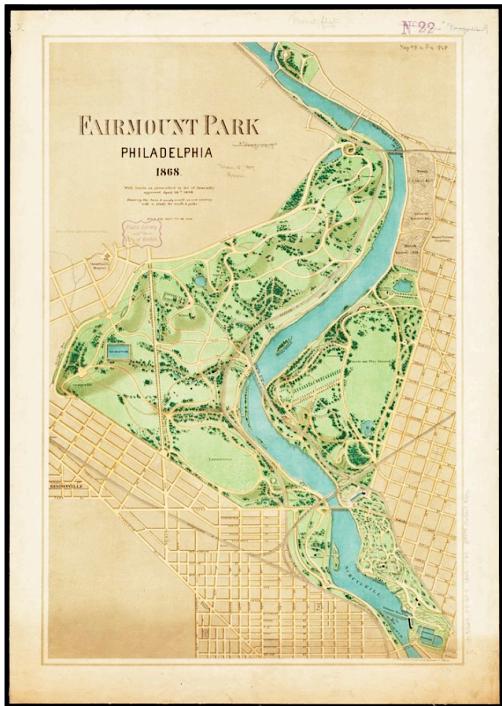
Sources Include but are not limited to⁷:

- [U.S. Census](#)
- [Canadian Census](#)
- [Foursquare Data](#)
- [Philadelphia vs. Toronto Web site](#)
- [Technical.ly Philly](#)
- [History of Toronto](#)
- [History of Philadelphia](#)
- [The Encyclopaedia of Philadelphia](#)
- [Toronto Neighbourhoods and Communities](#)
- [The Paris Review: America's First Female Map Maker](#)

⁷ This is the *short* list.

POSTSCRIPT

This:



.... Was the inspiration for my inquiry.