Visualizing COVID-19 Outbreaks at Long Term Care Facilities in Ontario

A Proposal to integrate the Open Database of Buildings and the Open Database of Healthcare Facilities with COVID-19 data

Shreeram Murali

KT Hobbs

Ngan Lyle

Sofia Bahmutsky

May 5, 2020

Partners:	Bruno St-Aubin
	Marian Radulescu
Instructor:	Scott Fazackerley

1   Introduction

Since its initial appearance in China on December 31, 2019, Coronavirus disease, or COVID-19, has evolved to a global pandemic — a novel disease to which there is little pre-existing immunity that becomes epidemic in many countries. At the time of writing, COVID-19 has spread to more than 180 countries (1). Worldwide, there are greater than 3 million confirmed cases and greater than 230,000 deaths (2). In Canada, there have been almost 60,000 confirmed cases with almost 18,000 cases and 1,300 deaths in Ontario alone (3). As the pandemic spread, it has become clear that seniors shoulder a disproportionate burden of disease. In fact, almost half of fatalities in Canada are related to outbreaks among seniors in long term care facilities (4). This proposal seeks to integrate a number of open source Statistics Canada databases to provide insight into the COVID-19 outbreak among long term care facilities in Ontario. Our goal is to create an interactive visualization of long term care facilities in Ontario as well as an article exploring the spread of COVID-19 among long term care facilities in Canada.

2   Resources

We will be using the following open databases from the Linkable Open Data Environment (LODE):

Open Database for Health Facilities (ODHF) data A comma-separated file of health facility information across Canada with missing address information.

Open Database of Buildings (ODB) data A spatial database of buildings across Canada. We will focus on data for Ontario.

Proximity data for transit, health care, pharmacies (time permitting) A comma-separated file of proximity metrics across Canada on the regional level. We will focus on data for Ontario.

In addition, we will be using the Ontario provincial health COVID data.

3 Proposed Project

3.1 Objectives

This project involves analyzing open source data and using interactive visualizations to produce an exploratory analysis tool for observing potential relationships within data. To build the required dataset, we will initially merge several sources: ODHF, ODB, proximity data, and COVID-19 data. Pursuits will focus on data for the province of Ontario. ODHF data will also require completing missing addresses, which we aim to accomplish by web scraping. Once the ODHF data is complete, it would be joined to ODB on location columns (latitude/longitude). Finally, proximity and COVID-19 datasets could also be joined on location columns.

Once data cleaning and integration is complete, it would be used to create interactive maps that depict location and severity of COVID-19 cases in senior care facilities in Ontario in addition to a time-lapse of the daily cases. Ideally, this would be accomplished using shape files from the ODB as well as GIS, and D3 to build a dashboard.

Finally, the created visualizations would be used for exploratory data analysis and observation of relationships within the data to gain insight about COVID-19 and the senior care facilities that are affected. Combined with proximity data, the visualizations could show connections between COVID-19 and other regional/spatial factors that may not have been considered previously.

There are two main objectives:

- To ensure quality and possible improvement of the open data sources (ODHF and COVID-19).

- To produce products showing that open data (including the databases that we are using from the LODE) are usable in complex analytical cases, resulting in high quality interactive visualizations and analyses.

3.2 Deliverables

- Complete and clean ODHF dataset with a focus on senior care facilities in Ontario.

- OBD with attributes linked to some building footprints, with a focus on healthcare facilities in Ontario.

- Visualizations integrating healthcare facilities, proximity metrics and COVID-19 outbreak data in Ontario.

- Research paper that includes exploratory analysis of patterns and/or connection between ODB, ODHF, proximity data, and COVID-19 data.
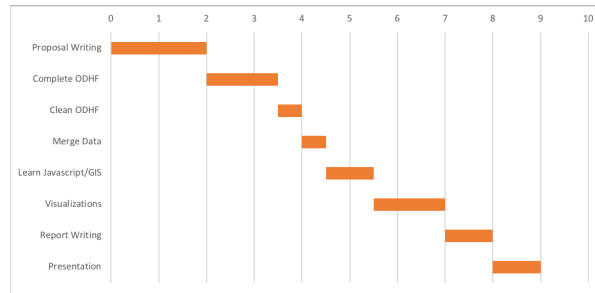
3.3 Schedule

Figure 1: Schedule for completion of tasks for the project in weeks.

3.3.1 Schedule Details

The project will begin as soon as permission is granted to proceed and will proceed according to the schedule shown in **Figure 1**. Details for selected tasks are shown in **Table 1**. The final report will be completed by the week of June 15th.

3.4 Limitations

The proposed project, including data preparation, analysis, visualization and report writing, is being undertaken on a compressed timeline and includes exposure to new tools and programming languages such as JavaScript and GIS. As a result, the deliverables will likely require assistance and further review before they are publishable.

Table 1: Details for Selected Tasks

Task	Details
Complete ODHF	**1.** Complete missing addresses
	2. Complete missing facility types
Clean ODHF	**1.** Ensure data types and structures are compatible between sources
	2. Resolve duplicate entries
Merge Data	**1.** Merge ODHF with ODB
	2. Merge with COVID-19 Ontario case data
Learn Javascript/GIS	Learn how to implement using D3
Visualizations	**1.** Interactive Choropleth Map of COVID-19 case numbers by health region in Ontario with time lapse
	2. Interactive map of long term care facilities in Ontario with COVID-19 metadata

4 Bibliography

1 John Hopkin's COVID-19 Dashboard

2 WHO arcGIS Dashboard

3 How Ontario is responding to COVID-19

4 Outbreaks at seniors' homes linked to almost half of COVID-19 deaths in Canada, Theresa Tam says

5 Confirmed positive cases of COVID-19 in Ontario