



Modeling and Visualization of the COVID-19 Outbreak in Ontario Statistics Canada

May 12, 2020

Sofia Bahmutsky
Kaitlyn Hobbs
Ngan Lyle (Presenter)
Shreeram Murali

Team



Ngan

KT

Sofia



Shreeram



Statistics
Canada

Statistique
Canada

Introduction

- Client
 - Statistics Canada
 - Bruno St-Aubin, Team Lead and GIS Developer
 - Marian Radulescu, Unit Head and Analyst
- Requirements
 - Assess the quality and possibly improve upon StatCan open data sources
 - Build a product that shows that StatCan open source data are useable in complex analytical cases

Purpose

- Analyze the COVID-19 Outbreak in Ontario
 - Spread in long term care (LTC) homes
 - In the media
 - Association between disease activity in different (Public Health Unit) PHU regions and StatCan proximity measures

e Star Edition
ANGE LOCATION

THE STAR

CANADA

POLITICS

WORLD

OPINION

LIFE

SPORTS

ENTERTAINMENT

BUSINESS

Federal Politics

Provincial Politics

Political Opinion

FEDERAL POLITICS

82% of Canada's COVID-19 deaths have been in long-term care, new data reveals

HEALTH

Coronavirus: Canadian military arrives at 5 Ontario long-term care homes struggling with COVID-19



BY JESSICA PATTON · GLOBAL NEWS

Posted April 24, 2020 9:15 am

Updated April 24, 2020 4:51 pm

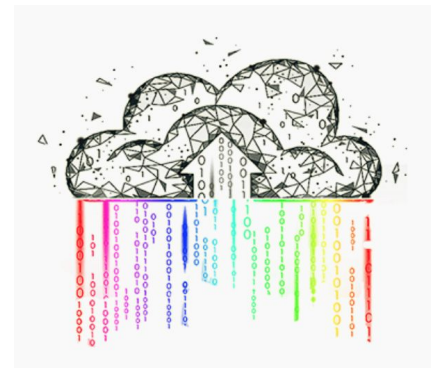


Objectives

1. Produce an inferential statistical model of factors that may be associated with the COVID-19 outbreak in LTC homes in Ontario
2. Produce an inferential statistical model of proximity factors that may be associated with COVID-19 outbreaks at the level of PHU regions in Ontario

Data

- StatCan Sources
 - Open Database for Health Facilities
 - Linkable Open Data Environment (LODE)
 - Data Exploration and Integration Lab (DEIL)
 - Proximity data for transit, health care, and pharmacies
- Other Sources
 - COVID-19 Data by Health Regions
 - Natural Resources Canada
 - Government of Ontario
 - Data about LTC homes in Ontario will be scraped from government websites

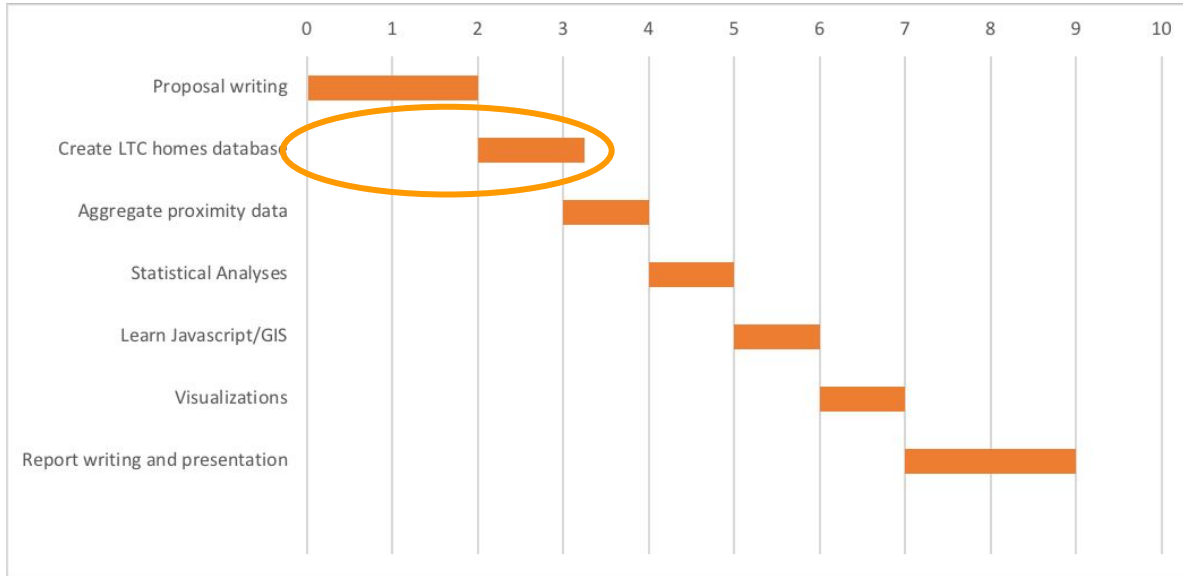


Tasks



1. Data Preparation
 - a. Create a database of LTC homes in Ontario
 - i. Data cleaning (Python)
 - ii. Webscraping (Python)
 - b. Aggregate proximity data to level of PHU regions
2. Statistical Analysis (R)
 - a. Principal component regression
 - b. Factor analysis
3. Visualization
 - a. Javascript, GIS and D3

Schedule



<https://trello.com/b/BOM8D6zv>



Deliverables

- A database of LTC homes in Ontario
 - Update StatCan ODHF
- A method to aggregate proximity data to the PHU region level
- Two inferential statistical models
 - Association between LTC home characteristics factors and COVID-19 outbreaks
 - Association between proximity factors and COVID-19 outbreaks among different Ontario PHU regions.
- Visualizations
- Report
 - Explore connections between the LTC homes data, the proximity data, and the COVID-19 data



Summary of Individual and Team Work Logs For Week 2

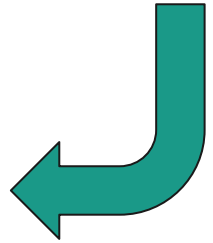
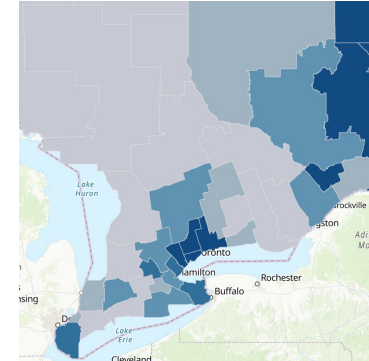
- Presentation of project to Prof Scott and TA's
- Proposal
 - Completed a draft and sent it to Statistics Canada on Tuesday
 - Reviewed with Bruno and Marian on Wednesday
 - Revised, resubmitted and received approval on Thursday
 - Submitted final proposal on Friday
- Began working in pairs on two data cleaning and web-scraping tasks
 - LTC characteristics data, Sofia and Ngan
 - LTC COVID-19 data, KT and Shreeram



Agenda for Week 3

- Web-scraping
- Data cleaning
- Merge data to form LTC homes database
- Develop a method to aggregate proximity data

- ## 2. Require arcGIS license





Thanks!