Modeling and Visualization of the COVID-19 Outbreak in Ontario

Statistics Canada

May 26, 2020

Sofia Bahmutsky (Presenter)
Kaitlyn Hobbs
Ngan Lyle
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

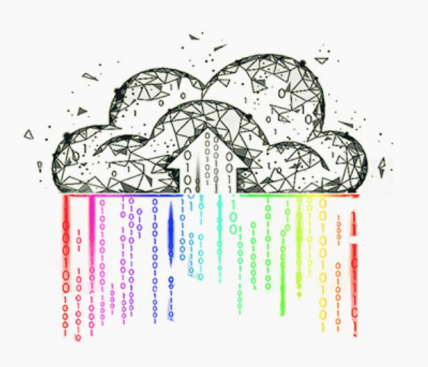
BY JESSICA PATTON - GLOBAL NEWS

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



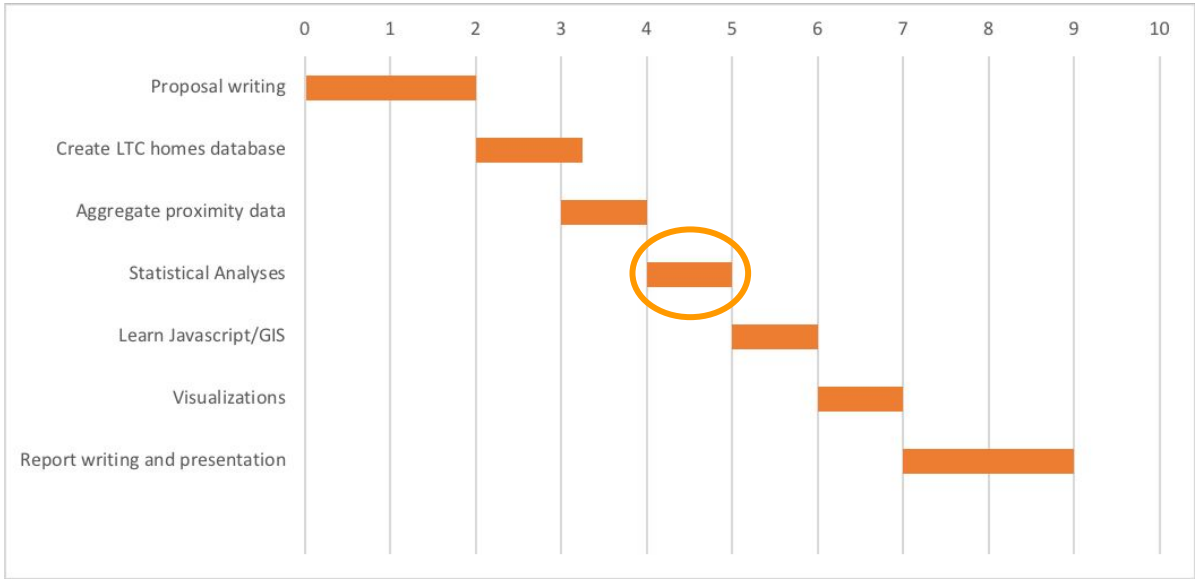
- StatCan Sources
  - Open Database for Health Facilities
    - Linkable Open Data Environment (LODE)
      - Data Exploration and Integration Lab (DEIL)
  - Proximity data for transit, health care, and pharmacies
- Other Sources
  - COVID-19 Data by Health Regions
    - Natural Resources Canada
   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

Tasks of Current Week

1. **Statistical Analysis**
    a. Research, reading of previous courses
    b. Correspondence with Jeff and John (profs)
    c. Complete analysis/modelling for LTC data and PHU data.

2. **If time allows, begin Javascript dashboard.**

Schedule



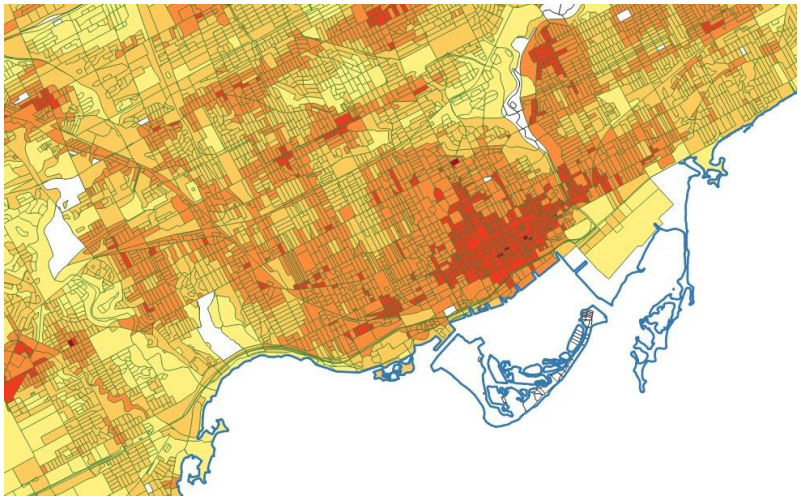
https://trello.com/b/BOM8D6zv

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 4

- Data cleaning and scraping is complete
- QGIS portion of project is complete
- Statistical Analysis is started



Agenda for Week 5

- Statistical Analysis
  - Need to determine optimal modelling method for our data
  - Want to incorporate the dimensionality-reduced or principle components to a clustering/classification scenario for LTC data, and a regression type analysis for PHU data.

Potential Limitations

1. We have mixed data types
2. Data need to be scaled, but this is causing some problems with analysis
3. Many approaches possible, need to figure out what is the best option
4. Limitations from getting help on the subject

Thanks!