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

June 23, 2020

Sofia Bahmutsky
KT Hobbs
Ngan Lyle
Shreeram Murali

Introduction

Team

Sofia

Ngan

KT

Shreeram

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

Background, Purpose & Objectives

Background

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

BARRIE | News

'It is not the right time to let your guard down," Region sees another jump in COVID cases

Kim Phillips Barrie.CTVNews.ca Web Producer
🐦 @ctvbarrienews | Contact

Published Monday, June 22, 2020 3:19PM EDT
Last Updated Monday, June 22, 2020 7:03PM EDT

Purpose

- Analyze the COVID-19 Outbreak in Ontario on two levels

    - Outbreak status in long term care (LTC) homes

    - Association between disease activity, health indicators and proximity measures in different Public Health Unit (PHU) regions

Objectives

1. To produce an inferential statistical model of factors that may be associated with the probability of COVID-19 outbreaks in different long-term care (LTC) homes in Ontario
2. To produce an inferential statistical model of the proximity and health factors that may be associated with COVID-19 disease activity at the level of the Public Health Unit (PHU) in Ontario
3. To produce an interactive webpage using QGIS and D3 to visualize the results from both PHU region level analysis and LTC homes level analysis
4. To leverage open data sources to produce meaningful exploratory analyses

Data & Methods

Data

- Statistics Canada Sources
  - Open Database for Health Facilities
    - Linkable Open Data Environment (LODE)
  - Proximity measures
  - Health indicators by PHU
- Other Sources
  - COVID-19 Data by PHU
    - Government of Ontario's Data Catalogue
  - Data about LTC homes in Ontario scraped from government websites

Data - PHU
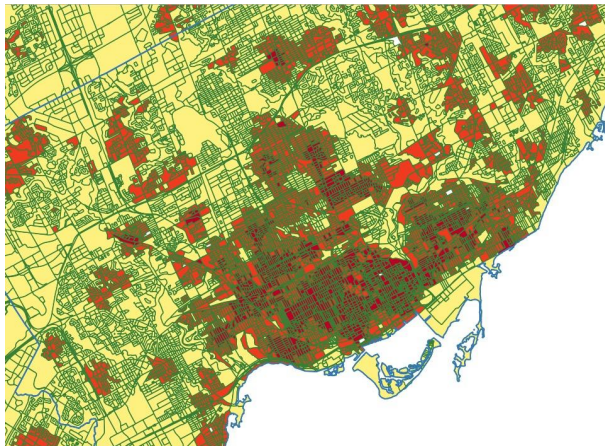
Proximity measures

```
Data columns (total 18 columns):
 #   Column
---  ------
 0   Unnamed: 0
 1   Row_ID
 2   Accurate_Episode_Date
 3   Case_Reported_Date
 4   Test_Reported_Date
 5   Specimen_Date
 6   Age_Group
 7   Client_Gender
 8   Case_AcquisitionInfo
 9   Outcome1
 10  Outbreak_Related
 11  Reporting_PHU
 12  Reporting_PHU_Address
 13  Reporting_PHU_City
 14  Reporting_PHU_Postal_Code
 15  Reporting_PHU_Website
 16  Reporting_PHU_Latitude
 17  Reporting_PHU_Longitude
```

Ontario COVID-19 Cases

```
Arthritis (15 years and over)
Asthma
Body mass index, adjusted self-reported, adult (18 years and over), obese
Body mass index, adjusted self-reported, adult (18 years and over), overweight
Body mass index, self-reported, youth (12 to 17 years old), overweight or obese
Breast milk feeding initiation
Chronic obstructive pulmonary disease (COPD; 35 years and over)
Current smoker, daily
Current smoker, daily or occasional
Diabetes
Exclusive breastfeeding, at least 6 months
Has a regular healthcare provider
Heavy drinking
High blood pressure
Influenza immunization in the past 12 months
Life satisfaction, satisfied or very satisfied
Mood disorder
Perceived health, fair or poor
Perceived health, very good or excellent
Perceived life stress, most days quite a bit or extremely stressful
Perceived mental health, fair or poor
Perceived mental health, very good or excellent
Physical activity, 150 minutes per week, adult (18 years and over)
Physical activity, average 60 minutes per day, youth (12 to 17 years old)
Sense of belonging to local community, somewhat strong or very strong
dtypes: float64(25)
```

Health Indicators

Data - LTC

Long-term care homes with an active outbreak [14]

An active COVID-19 outbreak indicates that the home has at least one lab confirmed case of COVID-19 (in resident or staff) and the local public health unit or the home has declared an outbreak.

The values represent the total cumulative number of residents that resided or staff that worked in the home, regardless if they were transferred to a hospital.

LTC Home	City	Beds	Confirmed Resident Cases	Resident Deaths	Confirmed Staff Cases
Albright Gardens Homes, Incorporated	Beamsville	231	0	0	0

Year 2020

Inspection Type	Inspection Report Date	Document
Complaints Inspection	Jan 17, 2020	Complaints Inspection Jan 17, 2020 - PDF (158 KB)

Year 2019

Inspection Type	Inspection Report Date	Document
Complaints Inspection	Nov 27, 2019	Complaints Inspection Nov 27, 2019 - PDF (150 KB)
Critical Incident Inspection	Nov 27, 2019	Critical Incident Inspection Nov 27, 2019 - PDF (145 KB)

Health Quality Ontario
Let's make our health system healthier

What is Health Quality | System Performance | Evidence to Improve Care
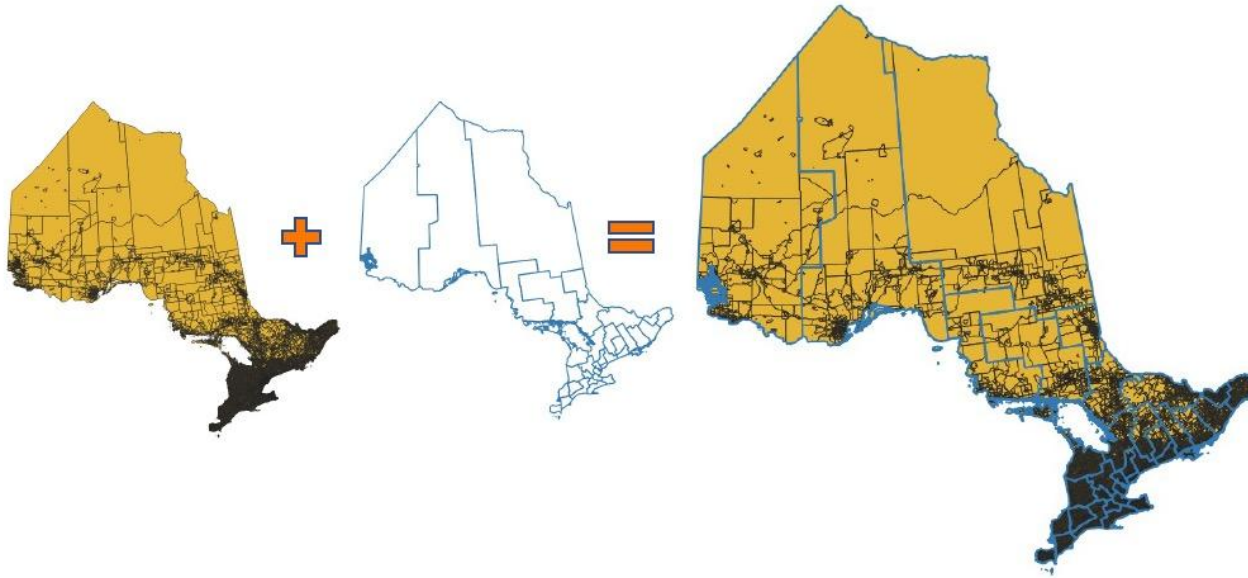
COVID-19: Get the latest update

SYSTEM PERFORMANCE

Home > System Performance > Long-Term Care Home Performance

Long-Term Care Home Performance in Ontario

These indicators provide data on wait times for admission to long-term care, long-term care home performance in Ontario.

12

QGIS data & methods

- Shapefiles from Stats Canada contained a lot of information which was extracted using QGIS.

- Using QGIS as a means for data wrangling.

- Aggregating layer shapefiles to classify each DB into its PHU.

Results

LTC Homes Analysis: Binary Logistic Regression

- Includes 615/625 homes without any missing data
- Binary outcome - outbreak status
- Predictors
  - Numeric
    - Number of beds
   - Quality (5)
    - Inspections (12)
  - Categorical (2)
  - Binary (4)

Variable	Estimate	Standard Error	P-value
Intercept	- 1.608	0.257	$4 * 10^{-10}$
Number of beds	+ 0.012	0.002	$1.41 * 10^{-10}$ ***
Total number complaints	+ 0.045	0.013	0.00065 ***
Total number non-complaints	- 0.025	0.011	0.025 *
Municipal home type	- 0.566	0.257	0.034 *
Non-profit home type	+ 0.167	0.267	0.437
For-profit home type (ref)	(ref)	(ref)	(ref)

LTC Homes Analysis

Log odds of an outbreak =

-1.608 + 0.012 * Number of beds

+ 0.045 * Total complaints

- 0.025 * Total non-complaints

- 0.566 * Municipal home type

McFadden R^2 = 0.14
p-value = 0

Outbreak status ordered by predicted probability of an outbreak
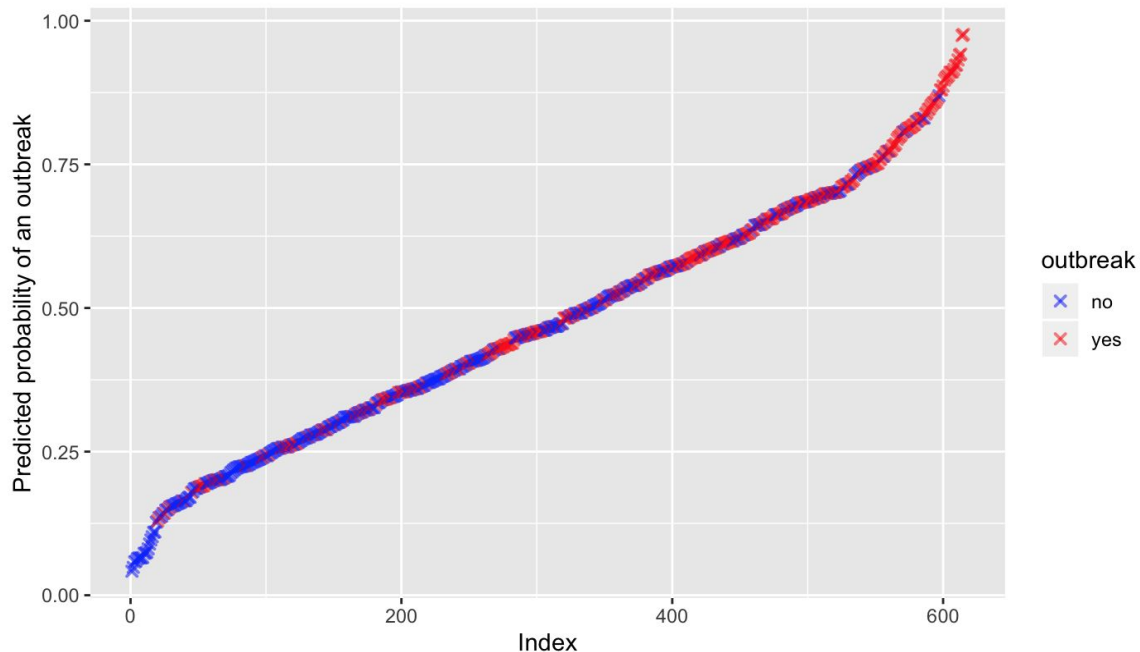
# PHU Analysis: Principal Components Analysis

Principal component 1 (PC1)

Proximity Variables (10)	Poor Health Variables (11)	Good Health Variables (9)
Positive ● Employment ● Pharmacy ● Childcare ● Healthcare ● Grocery ● Primary education ● Parks ● Transit Negative ● Secondary education ● Library	Positive ● Perceived life stress Negative ● Arthritis ● Obesity ● COPD ● Smoker ● Diabetes ● High blood pressure ● Mood disorder ● Perceived poor health ● Perceived poor mental health ● Asthma	Positive ● Overweight ● Immunization status ● Life satisfaction ● Perceived good health ● Perceived good mental health ● Physical activity > 150 min per week Negative ● Physical activity 60 min per day in youth ● Sense of community

PHU Analysis: PCA

PCA biplot

- Blue dots are PHUs
  - Similar PHUs cluster together
- Red arrows are predictors
  - Length and direction of the arrow indicates how strongly a predictor influences the PCs
  - Predictors that cluster together are correlated
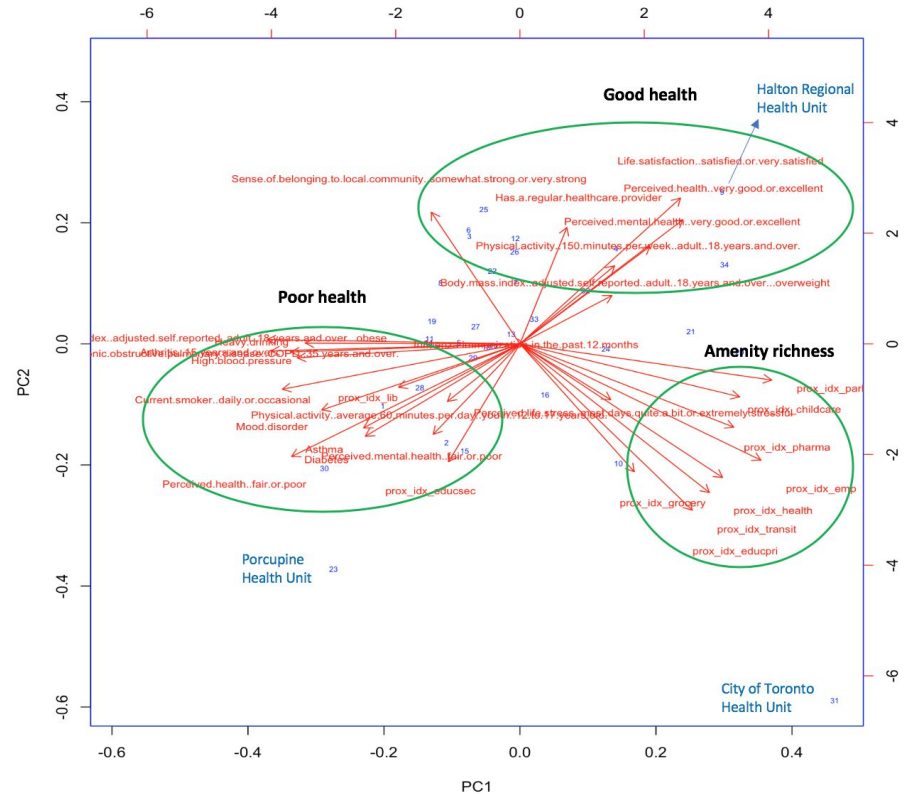
PHU Analysis: PC Regression
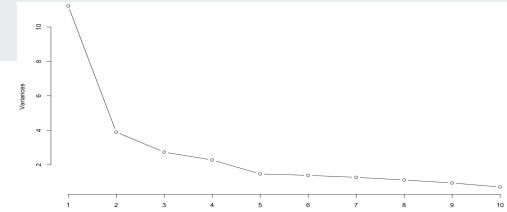
PCR on the proportion of cases

Principal Component	Estimate		Standard Error	P-value
Intercept	+	$1.378*10^{-3}$	$1.351*10^{-4}$	$8.62*10^{-11}$ ***
PC1	+	$2.258*10^{-4}$	$4.095*10^{-5}$	$7.71*10^{-6}$ ***
PC2	-	$1.340*10^{-4}$	$6.950*10^{-5}$	0.0644
PC3	+	$2.028*10^{-4}$	$8.299*10^{-5}$	0.0213 *

- Adjusted R2 = 0.5163
- P-value = 0.0001648

PCR on the proportion of fatalities

Principal Component	Estimate		Standard Error	P-value
Intercept	+	$1.034*10^{-4}$	$1.489*10^{-5}$	$1.84*10^{-7}$ ***
PC1	+	$1.907*10^{-5}$	$4.513*10^{-6}$	0.000243 ***

- Adjusted R2 = 0.3451
- P-value = 0.006226

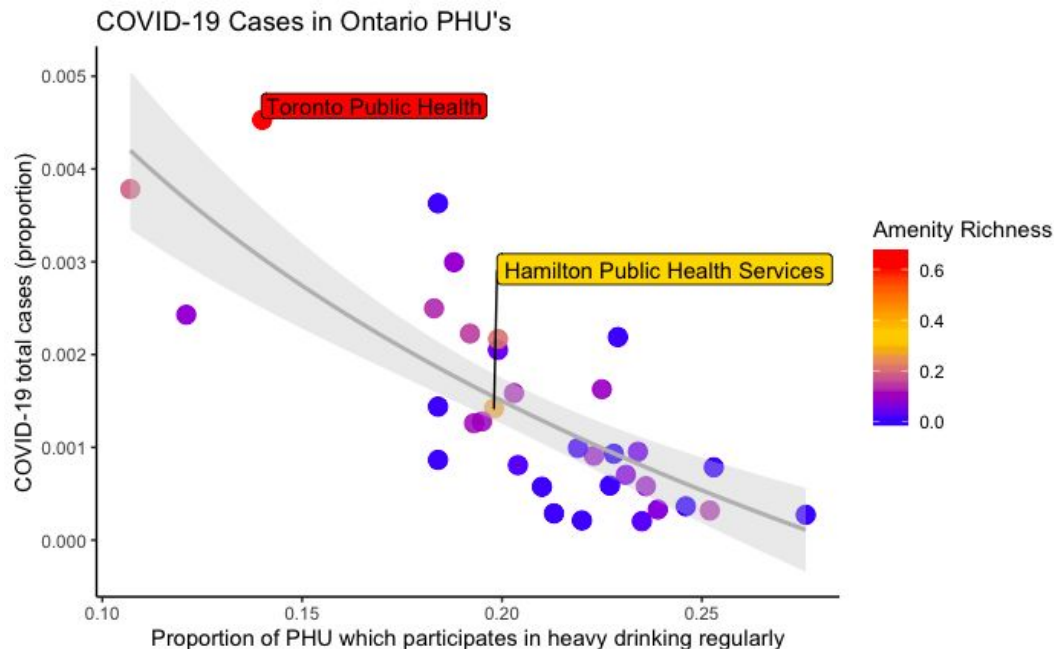**"Healthy and connected" increases your risk**

19

PHU Analysis

Agglomerative hierarchical clustering on health indicators and amenity score
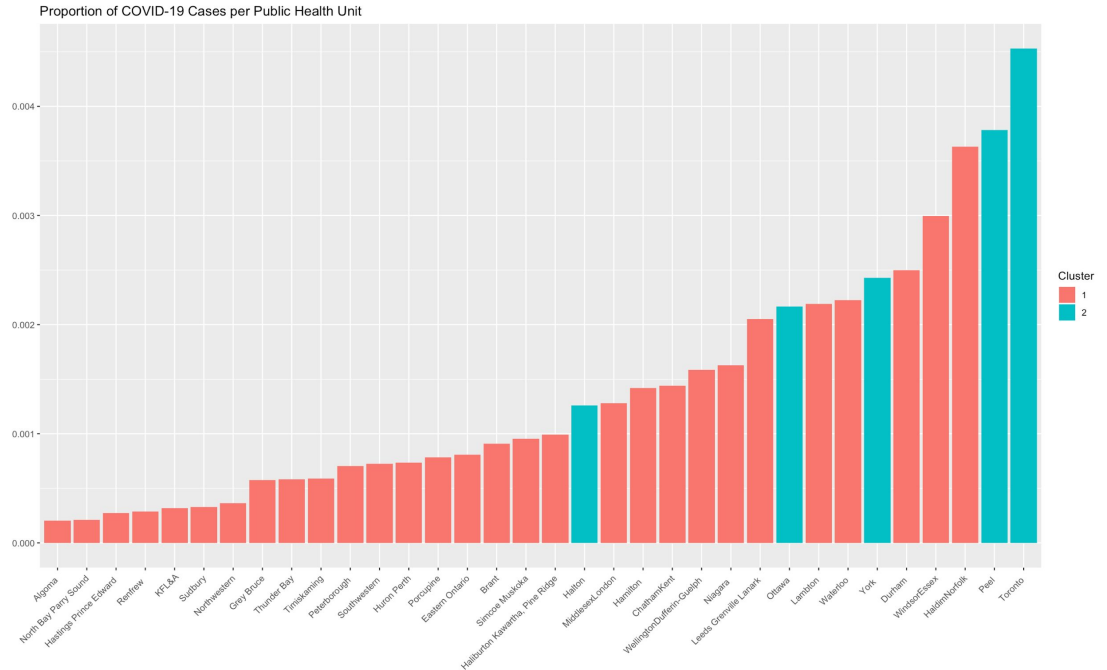
PHU Analysis

Supervised approach - COVID19 case proportion

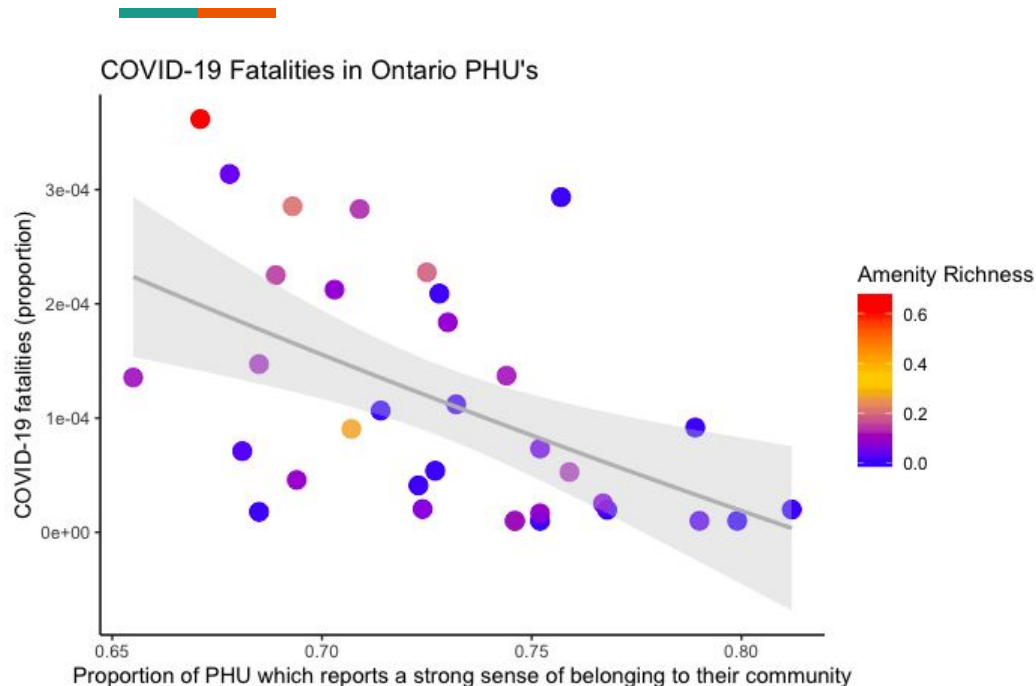COVID-19 Cases in Ontario PHU's

- LASSO and cross-validation for variable selection/reduction:
  1. Arthritis
  2. BMI - obese
  3. heavy drinking
  4. amenity density score.

- Beta regression model showed that only 'heavy drinking' was a significant predictor of COVID-19 cases

- Has a "protective" effect, ie PHU regions with high proportion of heavy drinkers are related to having lower COVID-19 cases.

loglog(COVID-19 case proportion) = -1.461 -1.519*(heavy drinking)

R-squared = 0.6

PHU Analysis

Supervised approach - COVID19 fatality proportion

COVID-19 Fatalities in Ontario PHU's

loglog(COVID-19 deaths proportion) = -1.206 -1.081*(strong sense of belonging)

R-squared = 0.557

- LASSO and cross-validation for variable selection/reduction:
  1. BMI - obese
  2. COPD
  3. heavy drinking
  4. strong sense of community belonging
  5. amenity density score.

- Beta regression model showed that only "strong sense of belonging" was a significant predictor of COVID-19 deaths

- Has a "protective" effect, ie PHU regions with more individuals having a strong sense of belonging are related to having lower COVID-19 cases.

Interactive Visualizations

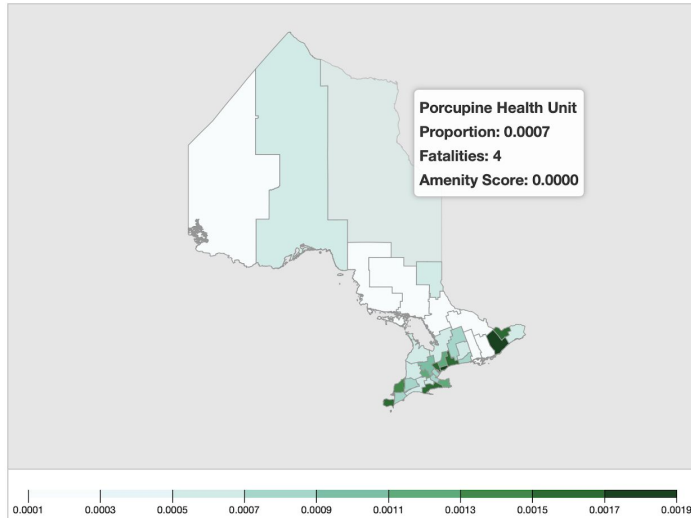https://ubco-mds-2019-labs.github.io/data-599-capstone-statistics-canada/kt/
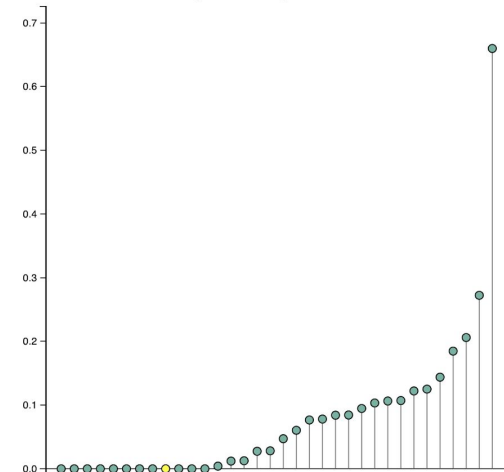
D3 Dashboard - Interactive Map

Ontario COVID-19 Map

Instructions

- Scroll in the map to zoom in and out.
- Click and drag the map to move.
- Hover over a public health unit or long-term care home (represented as points) to view details.
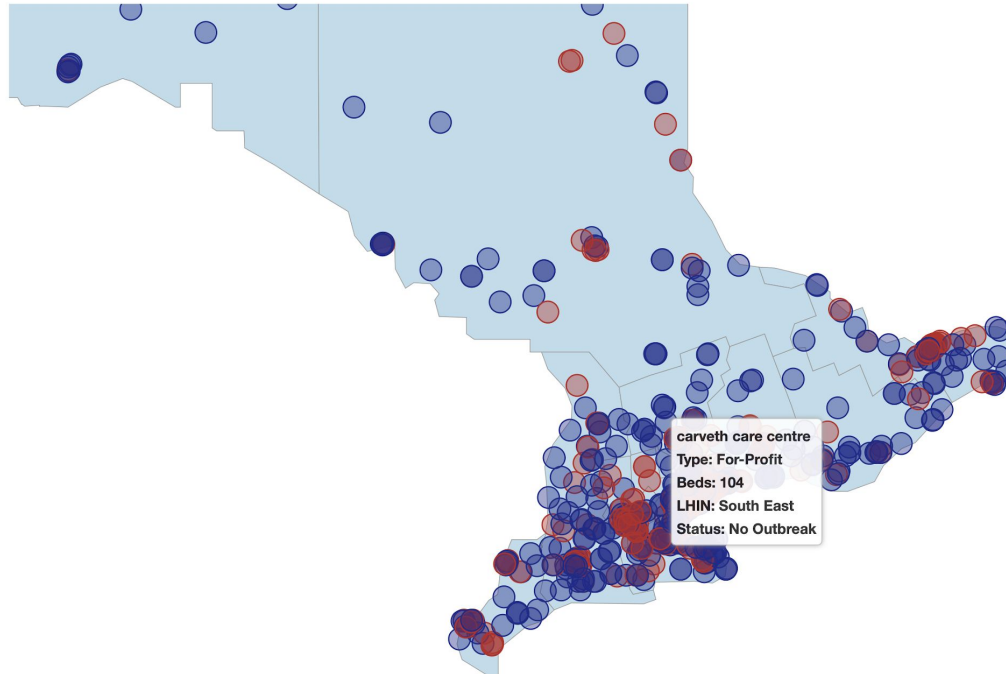

Porcupine Health Unit
Proportion: 0.0007
Fatalities: 4
Amenity Score: 0.0000

0.0001 0.0003 0.0005 0.0007 0.0009 0.0011 0.0013 0.0015 0.0017 0.0019

Amenity Richness by Public Health Unit

D3 Dashboard - Interactive Map

Long-Term Care Home Outbreaks in Ontario

carveth care centre
Type: For-Profit
Beds: 104
LHIN: South East
Status: No Outbreak

# Discussion, Limitations & Future Endeavours

Discussion

- No peer-reviewed publications regarding COVID-19 in Ontario found

- Researchers at UofT and Mount Sinai Hospital used same LTC data source

  - UofT - modeled temporal mortality trends

  - Mount Sinai - modeled extent of LTC outbreak

- No research to profile LTC homes or PHUs

Discussion

LTC home outbreak status was effectively modelled by logistic regression, with 4 predictor variables used for explanation of the response:
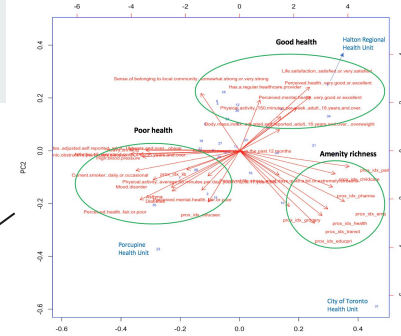
Number of beds, Total complaints, Total non-complaints, and Municipal home type

Recent news has highlighted the differences in quality of care which has been observed between municipal, for-profit, and non-profit homes in Ontario.
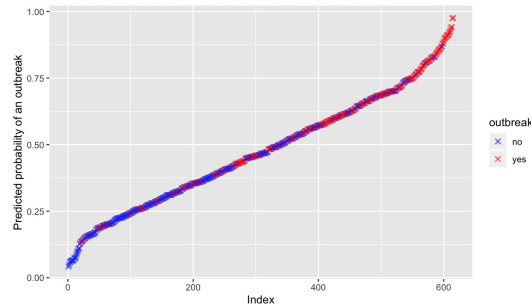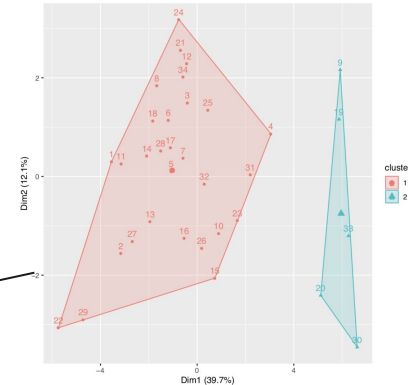
Discussion

- PCA analysis showed that there are 3 "drivers" when considering COVID-19 cases in a PHU: good health, poor health, and amenity density of the PHU.

Cluster Plot k = 2 – Agglomerative Complete Hierarchical Clustering

- Clustering analysis showed a clear divide which is best explained by "urban connectedness".

COVID-19 Fatalities in Ontario PHU's

- Supervised Method showed that there is a protective effect relating higher levels of heavy drinking to lower COVID-19 cases, and a relationship between higher sense of community to lower COVID-19 deaths.
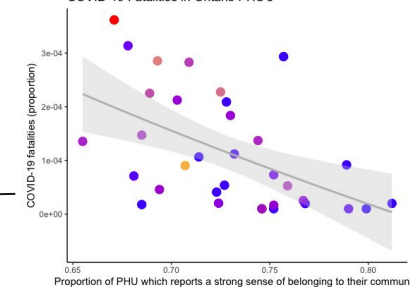
Limitations

- Limited time
  - Wrangling data from various sources
  - Learning curves associated with QGIS and JavaScript/D3
  - Temptation to do more!!!
- Data aggregation for COVID-19 cases
- Response variable for LTC analysis limited to binary

Future Work

- Socioeconomic analysis to better profile PHUs

- Further work on the dashboard

- Complementary statistical analysis

Thanks!

Additional Slides

LTC Third Scrape

- 66 active outbreaks

- 247 inactive outbreaks

- 313 total homes out of 647