Modeling and Visualization of the COVID-19 Outbreak in Canada

September 10, 2020

Ngan Lyle MD, MSDS

Background

- Started as a Capstone project at UBC Okanagan
  - Master of Data Science students
  - Mentored by Statistics Canada
  - Objectives for the Capstone project
    - Collect and organize data about the COVID-19 outbreak in Ontario
    - Show that open data can be used to produce meaningful analyses
- Current work
  - Refine and expand the statistical analysis
  - Replicate the analysis for other provinces

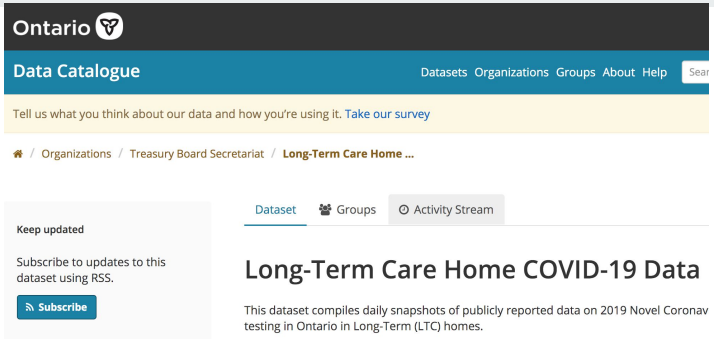
Aims

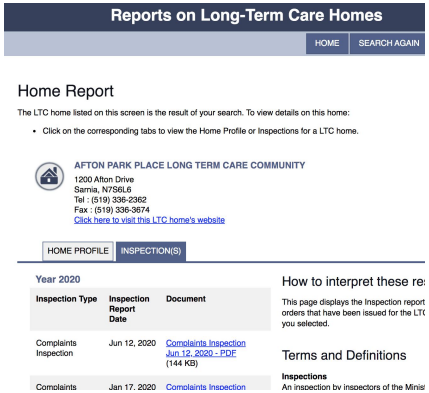
- Model the spread of COVID-19 among Long Term Care (LTC) homes
  - LTC homes analysis
- Model the spread of COVID-19 among health regions
  - Health regions analysis

LTC Homes Analysis for Ontario - Updated

LTC Homes - Data

- LTC homes Covid-19 data
  - Government of Ontario Data Catalogue
    - https://data.ontario.ca/dataset/long-term-care-home-covid-19-data
- LTC homes characteristics
  - Reports on Long-Term Care Homes
    - Ontario Ministry of Health and Long Term Care
    - http://publicreporting.ltchomes.net/en-ca/Default.aspx
  - Health Quality Ontario
    - Ontario Health
    - https://www.hqontario.ca/





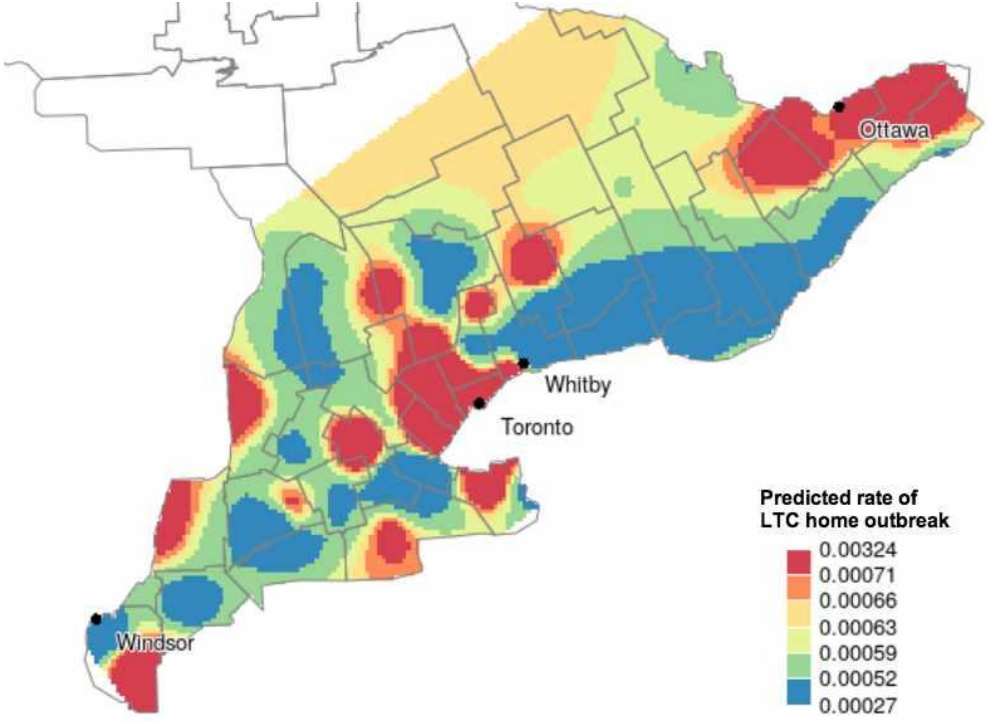
LTC Homes - Logistic Regression

- 615/625 homes without missing data
- Binary outcome - outbreak status
- Predictors (18)
  - Numeric (12)
    - Number of beds
    - Quality (5)
    - Inspections (6)
  - Categorical (2)
    - Operator type
    - Region
  - Binary (4)
    - Residents council
    - Family council
    - Accreditation
    - Short stay

Variable	Estimate	Standard Error	P-value
Intercept	- 3.57	0.59	$1.28 * 10^{-9}$
Number of beds	+ 0.36	0.04	$<2 * 10^{-16}$ ***
Total complaints/Number of beds	+ 2.90	0.96	0.0024 **
Total number non-complaints in the last 5 years	- 0.56	0.25	0.026 *
Municipal home type	- 0.61	0.26	0.020 *
Non-profit home type	+ 0.22	0.22	0.31
For-profit home type (ref)	(ref)	(ref)	(ref)

LTC Homes - Geostatistical Model

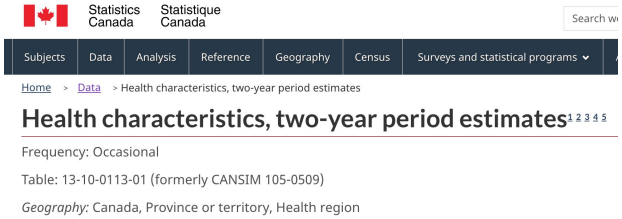
- Binomial model with individual effect
- Focus on LTC homes in Southern Ontario

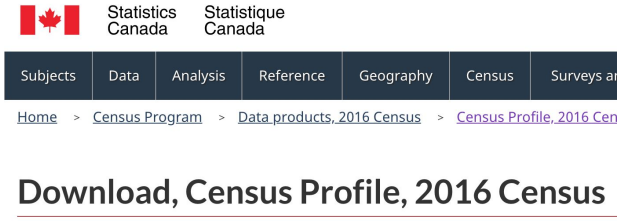


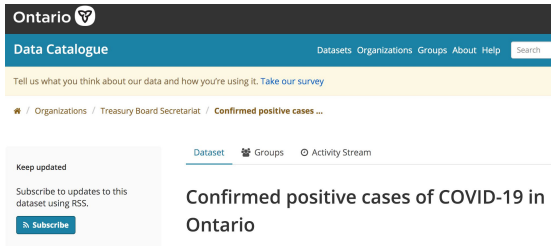
Health Regions Analysis for Ontario - Updated

Health Regions - Data

- Statistics Canada Data
  - Proximity data (10)
  - Health data (19)
    - Canadian Community Health Survey
  - **Census Profile data (26)**
    - **2016**
- Government of Ontario Data Catalogue
  - Confirmed positive cases of COVID-19 in Ontario
    - Cases
    - Fatalities







Health Regions - Principal Components Analysis (PCA)

PC1 42.7%

Variable type	Positive	Negative
Proximity	All 10 variables	None
Health	• Life satisfaction • Perceived good health • Perceived good mental health • Perceived life stress • Physical activity > 150 min per week **Good health factors**	• Medical problems • Mood disorder • Heavy drinking • Influenza immunization • Has a regular healthcare provider • Sense of belonging **Poor health factors**

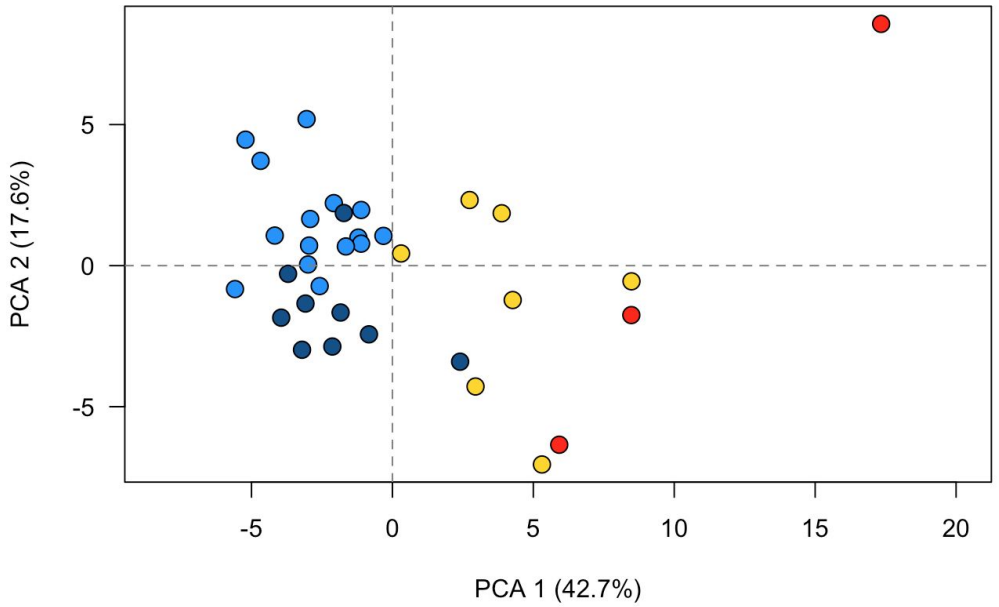
Census	• Population density
	• Demographic
	  ○ Younger age
	  ○ Single
	  ○ Higher income
	  ○ Higher level of education
	  ○ Working
	• Housing
	  ○ Renter
	  ○ Mover
	  ○ Non-suitable
	  ○ Larger household size
		• Older age
		• Married
		• Housing
		  ○ Detached
		  ○ Owner
		  ○ Non-mover
		  ○ Suitable
		• Driver

Health Regions - PCA

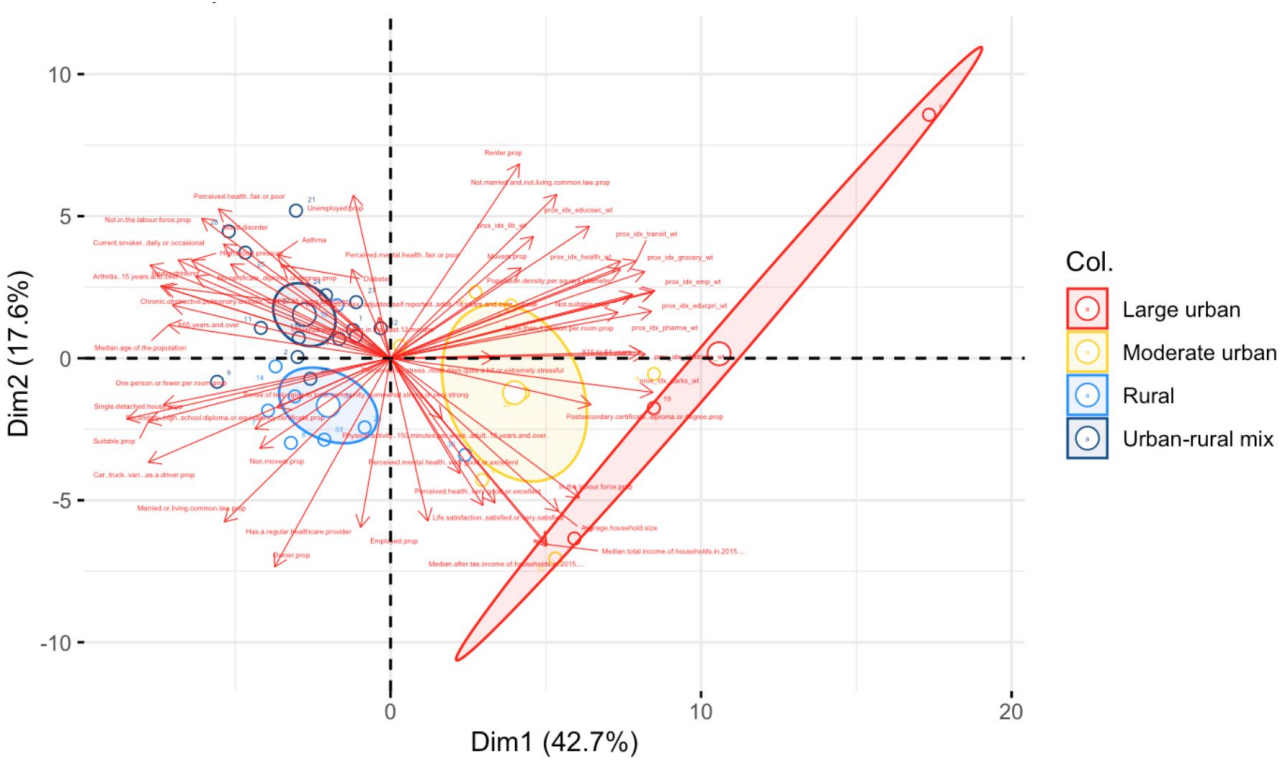
- Dots represent the health regions
- Similar health regions cluster together

More urban health regions;

1. **More connected**
2. **Young, healthy, mobile population**



Health Regions - PCA Biplot



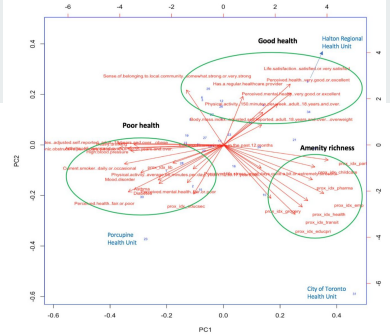
Health Regions - PC Regression

PCR on the proportion of cases

Principal Component	Estimate	Standard Error	P-value
Intercept	+ $1.43*10^{-3}$	$1.33*10^{-4}$	$2.84*10^{-11}$ ***
PC1	+ $1.21*10^{-4}$	$3.34*10^{-5}$	0.0012^{-6} **
PC3	+ $1.83*10^{-4}$	$7.19*10^{-5}$	0.017*
PC5	+ $1.21*10^{-4}$	$1.04*10^{-4}$	0.015 *

Regions that are "connected" with a "young and healthy" population are at increased risk

Replicate the Analysis for Other Provinces

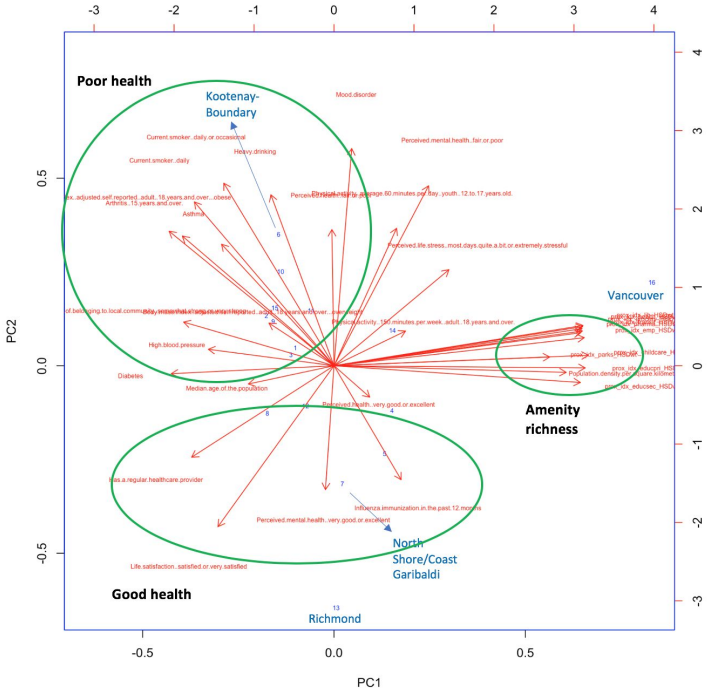


BC

- LTC Analysis
  - LTC Homes data rich and complex
    - > 100 predictors but many missing values

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-5.14887	0.90350	-5.699	1.21e-08	***
BEDS_TOTAL	0.37566	0.07486	5.018	5.21e-07	***
FAMILY_COUNCILTrue	-1.00846	0.49201	-2.050	0.0404	*

- Health Regions Analysis
  - Regions are not as clearly separate based on health factors
  - PC Regression showed a null result



Future Work and Challenges

- Ongoing work with the Ontario and British Columbia data
- Quebec
  - LTC data are not accessible
  - No central registry of LTC homes
  - Only homes with an active outbreak are identified on goverment websites
- Alberta
- Manitoba

Thanks!

- **UBCO team**
  - Prof. John Braun
    - KT Hobbs (BC/AB data)
    - Shahrukh Alvi (QC/MB data)
- **University of Toronto team**
  - Prof. Patrick Brown
    - Maggie Ma (Spatial Statistics)

LTC Homes - Logistic Regression

Log odds of an outbreak =

- 3.57

+ 0.36 * Number of beds

+ 2.90 * Total complaints (normalized)

- 0.56 * Total number non-complaints (5 years)

- 0.61 * Municipal home type