THE ONE

Lin Han

August 9th, 2020

COVID – 19 Fatal Probability Calculator

-- Data Solution to Reveal Deeper Reasons Behind COVID – 19 Fatal Cases in Toronto

THE ONE

Lin Han - Schulich New Graduate

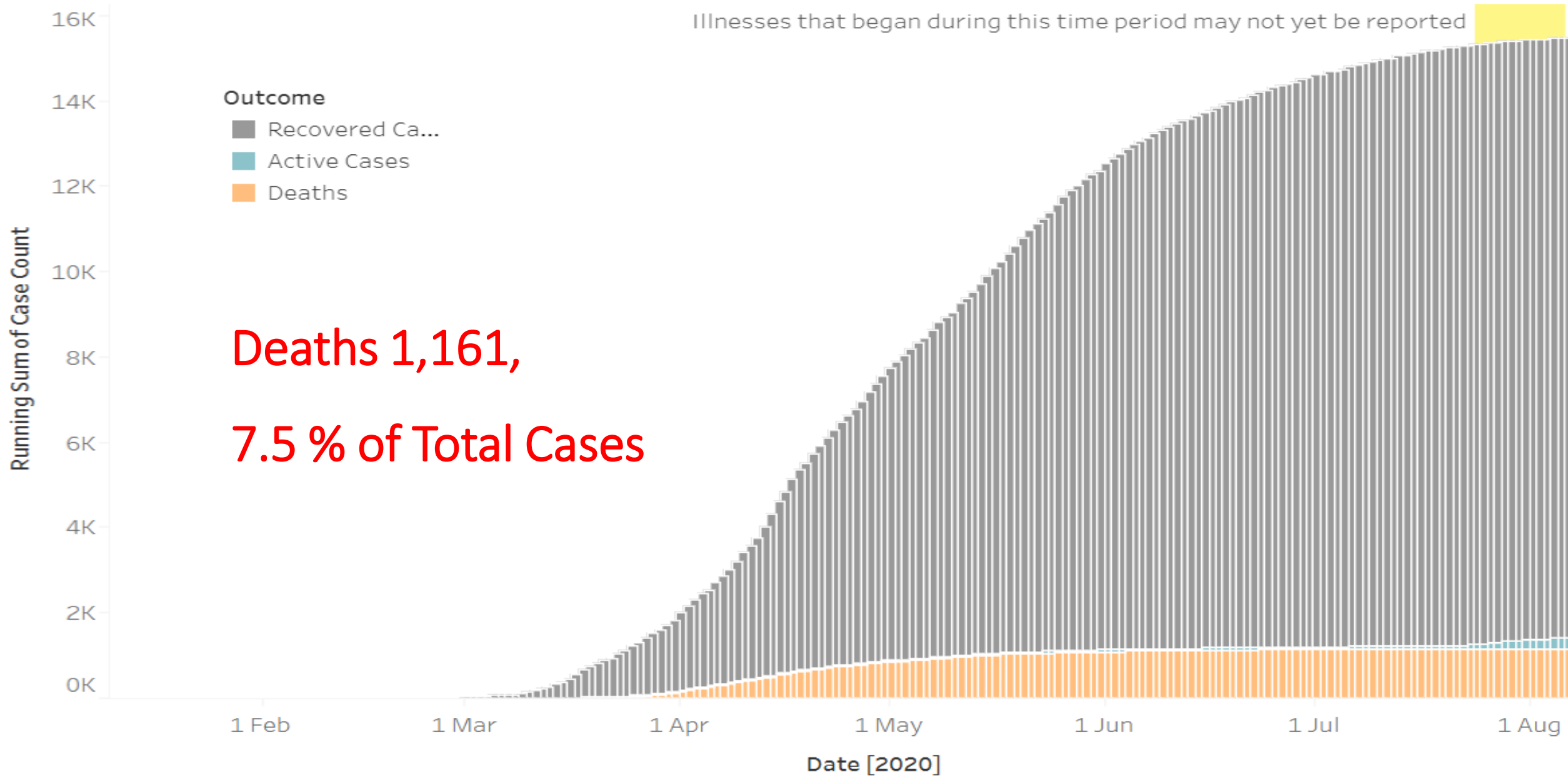# Agenda

I.   Project Background

II.   COVID – 19 Variables Analysis

III.   Fatal Probability Calculator

IV.   Calculation Examples

V.   Q & A
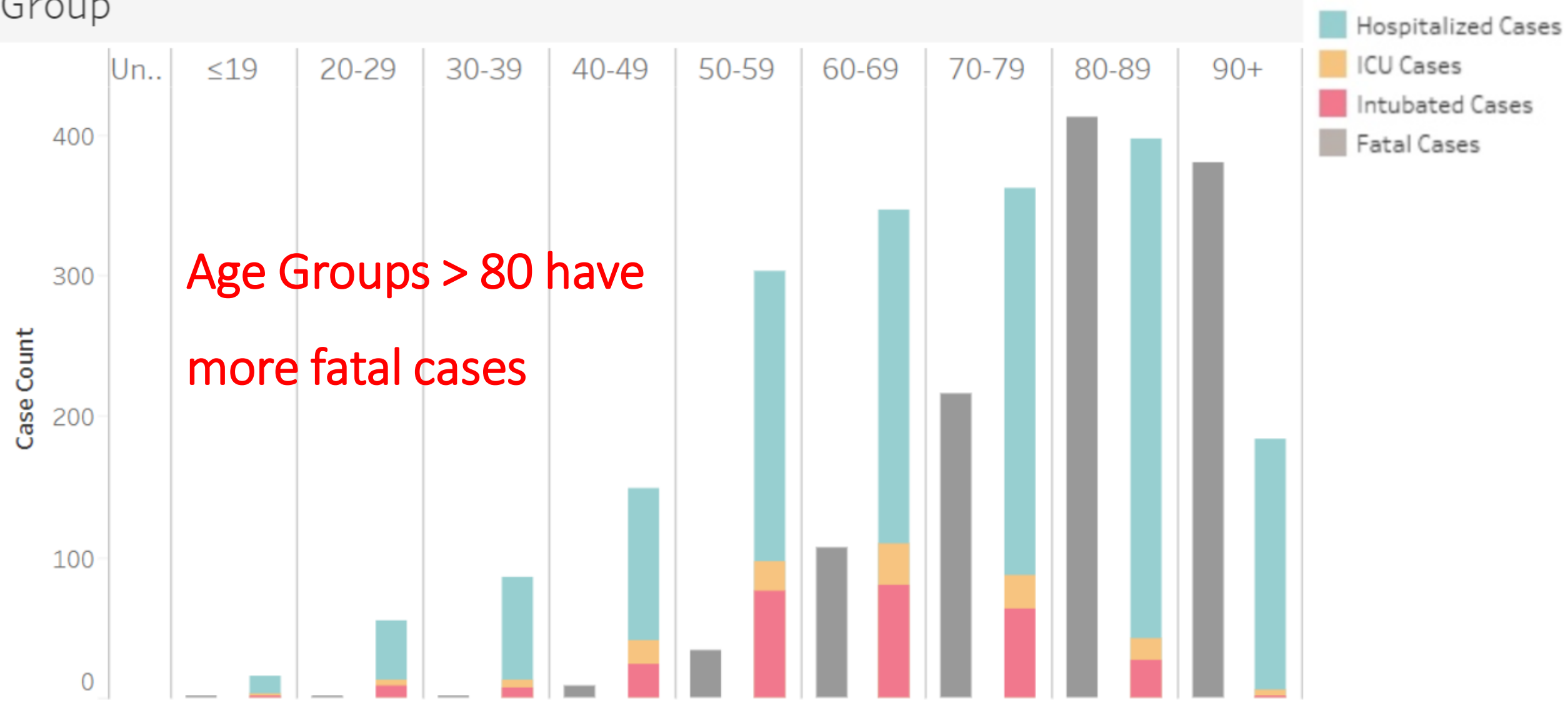
Cumulative Cases by Episode Date and Outcome

Episode Date
Reported Date

Illnesses that began during this time period may not yet be reported

Outcome
- Recovered Ca...
- Active Cases
- Deaths

Deaths 1,161,

7.5 % of Total Cases

Running Sum of Case Count

Date [2020]

City of Toronto: https://www.toronto.ca/home/covid-19/covid-19-latest-city-of-toronto-news/covid-19-status-of-cases-in-toronto/
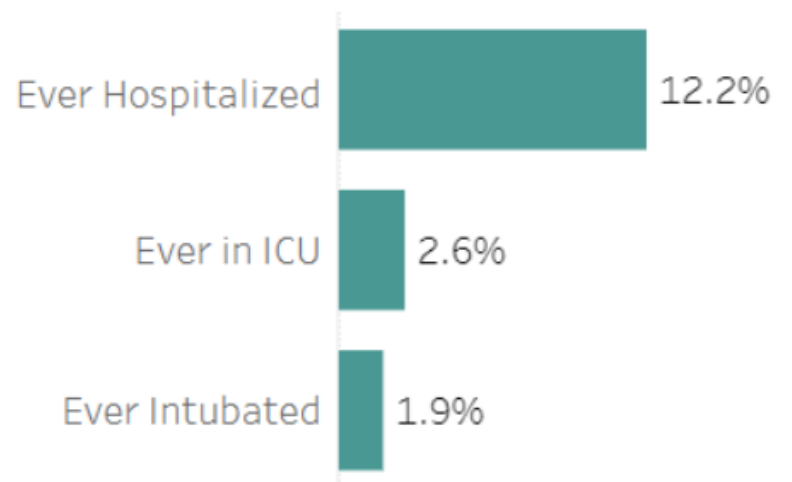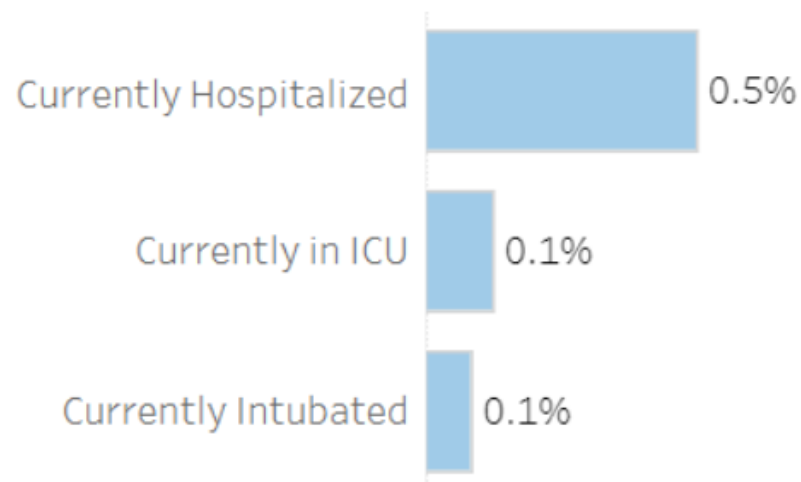
Number of COVID-19 Cases that Ever Resulted in Hospitalization, Intensive Care Unit (ICU) Admission, Intubation, and Deaths, by Age Group

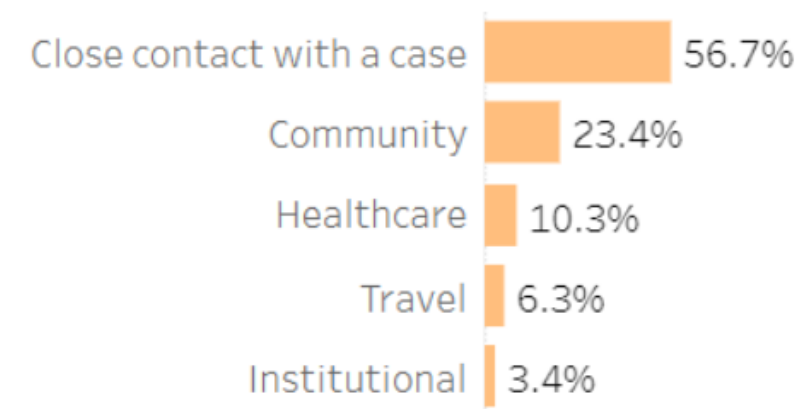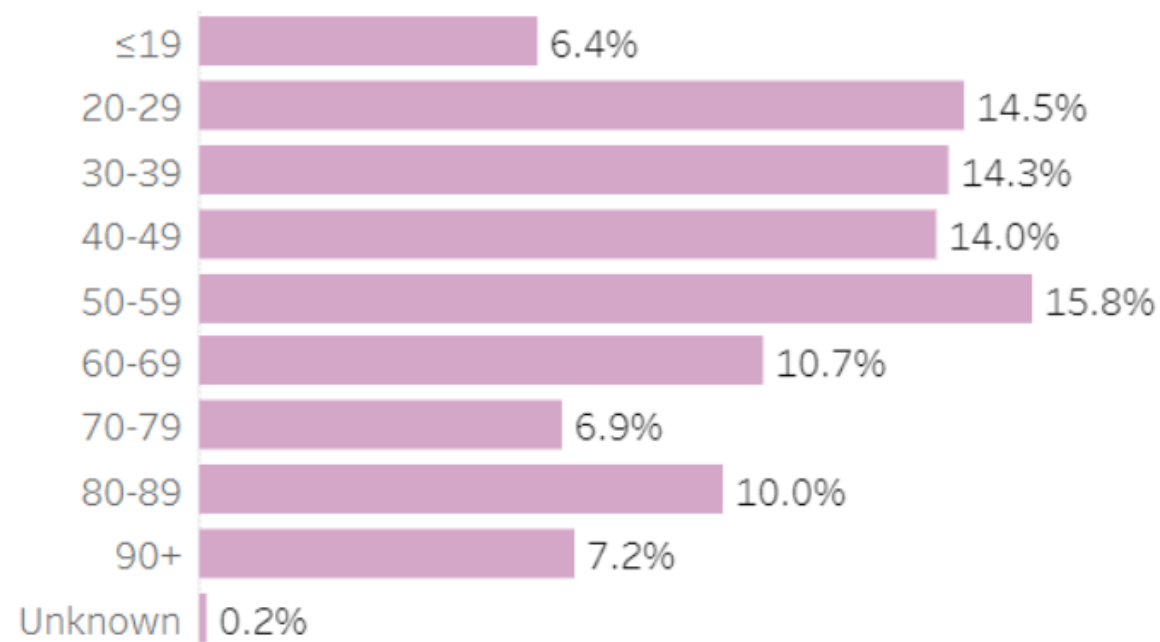Age Groups > 80 have more fatal cases

City of Toronto: https://www.toronto.ca/home/covid-19/covid-19-latest-city-of-toronto-news/covid-19-status-of-cases-in-toronto/

Percent Ever Hospitalized

Ever Hospitalized	12.2%
Ever in ICU	2.6%
Ever Intubated	1.9%

Currently Hospitalized

Currently Hospitalized	0.5%
Currently in ICU	0.1%
Currently Intubated	0.1%

Source of Infection - Sporadic Cases ‡

Close contact with a case	56.7%
Community	23.4%
Healthcare	10.3%
Travel	6.3%
Institutional	3.4%

Cases by Age Group

≤19	6.4%
20-29	14.5%
30-39	14.3%
40-49	14.0%
50-59	15.8%
60-69	10.7%
70-79	6.9%
80-89	10.0%
90+	7.2%
Unknown	0.2%

Cases by Gender

Female	52.9%
Male	45.5%
Unknown	1.6%
Other	<0.1%
Transgender	<0.1%
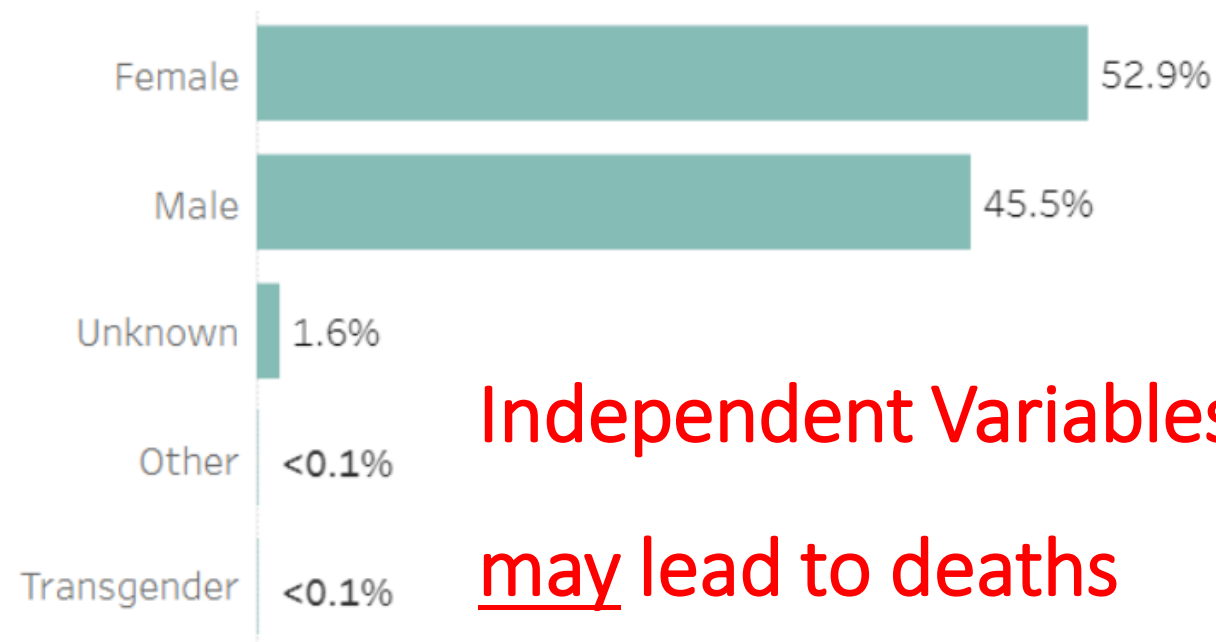
Independent Variables

may lead to deaths

City of Toronto: https://www.toronto.ca/home/covid-19/covid-19-latest-city-of-toronto-news/Covid-19-status-of-cases-in-toronto/

Dig Deeper into Factors Cause Deaths

Removed 5 highly correlated variables:

'_id', 'currently_intubated', 'currently_in_icu', 'ever_in_icu', 'ever_intubated'.

	_id	outbreak_associated	confirmed	currently_hospitalized	currently_in_icu	currently_intubated	ever_hospitalized	ever_in_icu
_id	1	-0.26	-0.027	0.025	0.01	0.0071	-0.1	-0.048
outbreak_associated	-0.26	1	0.18	-0.0039	-0.014	-0.01	0.036	-0.062
confirmed	-0.027	0.18	1	0.017	0.011	0.0087	0.075	0.038
currently_hospitalized	0.025	-0.0039	0.017	1	0.52	0.42	0.2	0.19
currently_in_icu	0.01	-0.014	0.011	0.52	1	0.82	0.1	0.23
currently_intubated	0.0071	-0.01	0.0087	0.42	0.82	1	0.083	0.19
ever_hospitalized	-0.1	0.036	0.075	0.2	0.1	0.083	1	0.44
ever_in_icu	-0.048	-0.062	0.038	0.19	0.23	0.19	0.44	1
ever_intubated	-0.056	-0.058	0.033	0.19	0.24	0.23	0.37	0.81
duration	-0.0021	-0.15	-0.064	-0.014	-0.0053	-0.0025	-0.02	0.013
episode_month	0.71	-0.14	0.018	0.04	0.019	0.014	-0.1	-0.069

Toronto Open Data 'COVID-19 CASES IN TORONTO' : https://open.toronto.ca/dataset/covid-19-cases-in-toronto/

Logistic Regression Model

-> 15 significant variables

'Age > 70' are more likely to die compared with 'Age < 50' because of COVID – 19.

```
                          Results: Logit
==============================================================================
Model:                  Logit            Pseudo R-squared:     inf
Dependent Variable:     outcome_fatal    AIC:                  inf
Date:                   2020-08-09 12:39 BIC:                  inf
No. Observations:       10364            Log-Likelihood:       -inf
Df Model:               14               LL-Null:              0.0000
Df Residuals:           10349            LLR p-value:          1.0000
Converged:              1.0000           Scale:                1.0000
No. Iterations:         11.0000
------------------------------------------------------------------------------
                     Coef.    Std.Err.     z      P>|z|   [0.025   0.975]
------------------------------------------------------------------------------
confirmed           -1.1483   0.2032   -5.6516   0.0000  -1.5465  -0.7501
ever_hospitalized    1.7442   0.1012   17.2339   0.0000   1.5459   1.9426
duration            -0.0285   0.0073   -3.9128   0.0001  -0.0428  -0.0142
episode_month       -0.6323   0.0432  -14.6424   0.0000  -0.7169  -0.5477
age_30 to 39 years  -2.8325   1.0115   -2.8004   0.0051  -4.8150  -0.8501
age_40 to 49 years  -1.2285   0.4738   -2.5926   0.0095  -2.1572  -0.2998
age_60 to 69 years   1.2970   0.2041    6.3538   0.0000   0.8969   1.6970
age_70 to 79 years   2.4162   0.1890   12.7846   0.0000   2.0458   2.7866
age_80 to 89 years   2.8525   0.1819   15.6858   0.0000   2.4961   3.2089
age_90 and older     3.5033   0.1886   18.5792   0.0000   3.1337   3.8728
source_close contact -1.0371  0.1651   -6.2815   0.0000  -1.3607  -0.7135
source_community    -0.8230   0.1910   -4.3077   0.0000  -1.1974  -0.4485
source_healthcare   -0.6684   0.2265   -2.9514   0.0032  -1.1123  -0.2245
source_travel       -1.4308   0.3209   -4.4593   0.0000  -2.0597  -0.8019
gender_female       -0.5194   0.0906   -5.7314   0.0000  -0.6970  -0.3418
==============================================================================
```

Variables Importance Rating

Top variables correlated to deaths: 'Duration of episode_date & reported_date',

'Age 80 and older' , 'Episode month'
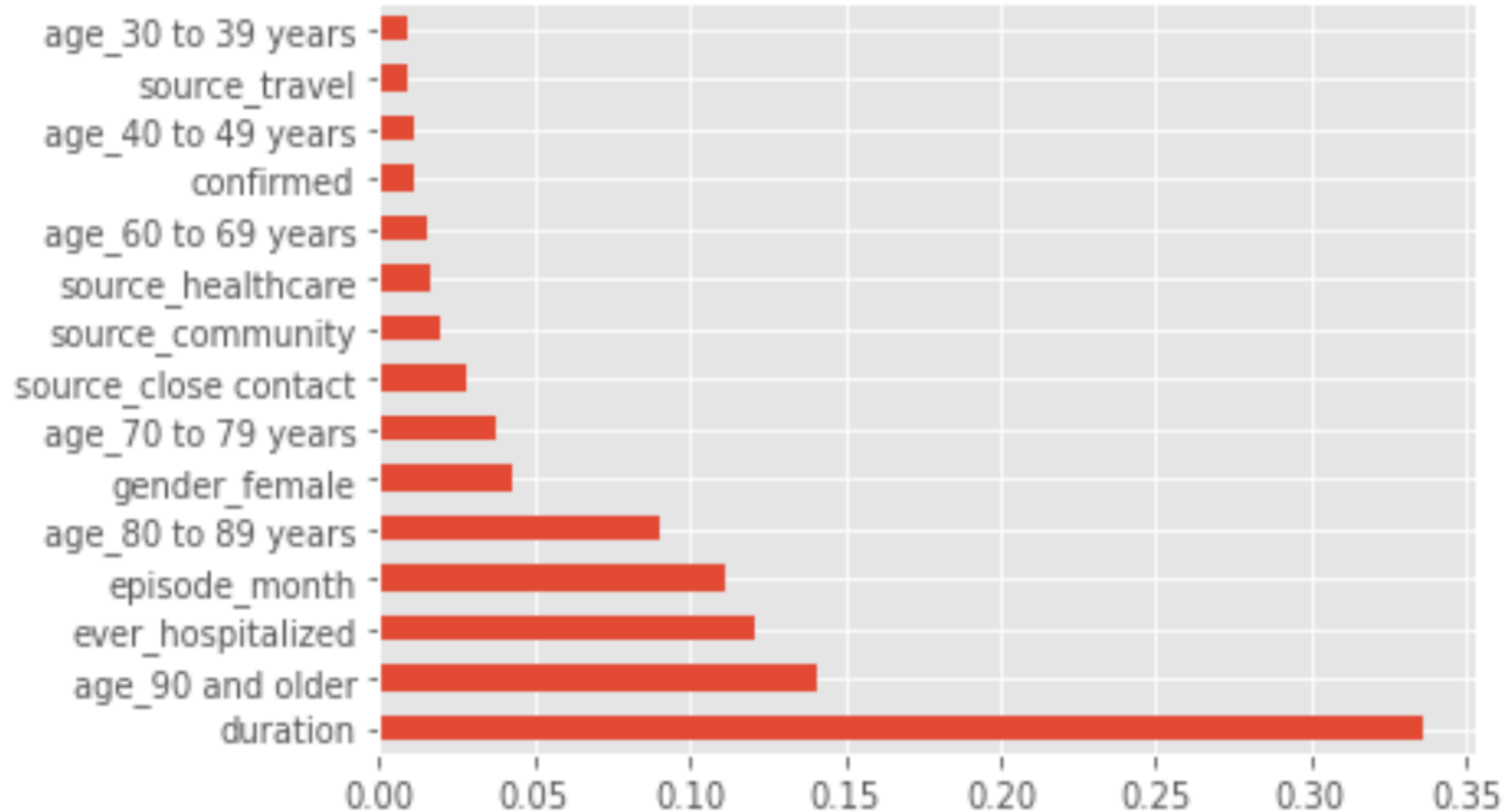
Fatal Probability Calculator

COVID – 19 Fatal Probability (P)

= 1 / 1 + exp{- [confirmed *(-1.148) + ever_hospitalized * (1.744) + duration *

(-0.029) + episode_month * (-0.632) + age_30 to 39 years * (-2.833) + age_40 to 49

years * (-1.229) + age_60 to 69 years * (1.297) + age_70 to 79 years * (2.416) +

age_80 to 89 years * (2.852) + age_90 and older * (3.503) + source_close contact *

(-1.037) + source_community * (-0.823) + source_healthcare * (-0.668) +

source_travel * (-1.431) + gender_female * (-0.519)]}

(Except for variables 'duration' and 'episode_month' are integer data type, other variables are dummy variables: 0 or 1.)

Fatal Probability Calculation – Example 1

Given a confirmed COVID – 19 patient A, who has in hospital, and duration is 10 days, episode month is 7, age 35, and source is community, gender female.

COVID – 19 Fatal Probability (P)

= 1 / 1 + exp{- [1 *(-1.148) + 1 * (1.744) + 10 * (-0.029) + 7 * (-0.632) + 1 * (-2.833) + 0 * (-1.229) + 0 * (1.297) + 0 * (2.416) + 0 * (2.852) + 0 * (3.503) + 0 * (-1.037) + 1 * (-0.823) + 0 * (-0.668) + 0 * (-1.431) + 1 * (-0.519)]}

= 1 / 1 +exp[-(-8.293)] = 1 / 3996.80 = 0.025 % -> will not likely die

Fatal Probability Calculation – Example 2

Given a confirmed COVID – 19 patient B, who has in hospital, and duration is 30 days, episode month is 2, age 90, and source is close contact, gender male.

COVID – 19 Fatal Probability (P)

= 1 / 1 + exp{- [1 * (-1.148) + 1 * (1.744) + 30 * (-0.029) + 2 * (-0.632) + 0 * (-2.833) + 0 * (-1.229) + 0 * (1.297) + 0 * (2.416) + 0 * (2.852) + 1 * (3.503) + 1 * (-1.037) + 0 * (-0.823) + 0 * (-0.668) + 0 * (-1.431) + 0 * (-0.519)]}

= 1 / 1 +exp[-(0.928)] = 1 / 1.3953 = 71.67 % -> will likely die

Reference

1. Picture source:
http://www.pentictonherald.ca/news/national_news/article_a29f7353-8eb4-5fd5-a7e5-5d31d16d43b5.html

2. City of Toronto, COVID-19: Status of Cases in Toronto:
https://www.toronto.ca/home/covid-19/covid-19-latest-city-of-toronto-news/covid-19-status-of-cases-in-toronto/

3. Toronto Open Data 'COVID-19 CASES IN TORONTO' :
https://open.toronto.ca/dataset/covid-19-cases-in-toronto/

4. Probability Calculation Using Logistic Regression
https://docs.tibco.com/pub/sfire-dsc/6.5.0/doc/html/TIB_sfire-dsc_user-guide/GUID-C4D05ED0-3392-4407-B62A-7D29B26DC566.html