Intro (rough)

The novel corona virus of 2019 dubbed Covid-19 has profoundly altered life on earth in 2020. In addition to the human sickness and death associated with a pandemic, widespread economic shutdown has been implemented across the globe in an effort to limit the spread of the virus. As students, our lives have been directly impacted as international travel has been vastly reduced, classes moved to online delivery, and personal connections limited. This unique situation provides us with an importunity to study a problem that is hugely relevant to our day-to-day lives and where results may become immediately useful in the fight to control this disease.

As a result of these unique circumstances, we have decided to mine Canadian Covid-19 patient data in an effort to determine which demographics are most at risk of hospitalization and death and if these risk demographics change depending on geographic location. Utilizing data provided by federal and provincial government we intend to use a variation of the FP-growth algorithm to determine the most at-risk demographics around the country. Patient data generally includes age, sex, health status, and exposure type which we will mine to determine frequent attributes and hopefully draw some conclusion with regards to high-risk demographics. In addition, in comparing our results province-by-province we will determine if risk has any region-dependency.

Our contributions will be two-fold. First, we will examine high-risk demographics and how they change depending on province. While risk demographics associated with Covid-19 have been studied regionally in Canada already [see source in related work about BC, ON, QC], due to the rapidly evolving nature of an on-going pandemic experiencing a second wave of infections, our project will provide more results on a larger set of cases over more provinces. As British Columbia, Ontario, and Quebec contained the vast majority of the initial infections in spring of 2020, most research has been conducted on these specific provinces. At the time of writing (November, 2020) Alberta, Saskatchewan, and Manitoba are all experiencing significant outbreaks which will add a much larger sample to our mining. Second, We will alter the classic FP-growth algorithm to suit our particular needs efficiently. Removing a second pass from the algorithm will reduce computation time and allow for quick data mining as the Covid-19 patient data available continues to grow throughout this pandemic.