

# BC Chronic Disease Capstone Proposal

Jennifer Hoang, Jessie Wong, Mahmoodur Rahman, Shengting (Irene) Yan

May 10, 2022

## Contents

Executive Summary . . . . .	1
Introduction . . . . .	1
Data Science Techniques . . . . .	1
Timeline . . . . .	2
Conclusion . . . . .	2

## Executive Summary

The Chronic Disease Registry is a data product that captures information about the rate of new and persistent cases of chronic diseases in each CHSA (community health service area), as well as demographic information such as sex and age standardized rates. In this project we aim to create an interactive dashboard that will allow users of all technical expertise to explore and visualize spatiotemporal information of the diseases in the data, and to develop an analysis pipeline that will describe the spatiotemporal trends in the data. This proposal will outline the approach we will take to tackle this problem and achieve the project goals.

## Introduction

Millions of people in BC live with a chronic disease, so it's important to understand and interpret the distribution of disease prevalence throughout the province for a variety of reasons. We may want to know how to best allocate healthcare resources, or to identify if a specific region is experiencing rapid growth of a disease.

To visualize the spatial and temporal trends of disease prevalence in the province, we will build an interactive dashboard that will allow users to compare the incidence rates of various diseases in one specific health region over time, as well as to compare how the incidence and prevalence rates of one disease has varied across several health regions over time. We also plan to create an information page in the dashboard containing descriptions of variables and diseases to increase usability for non-technical users. Lastly, we will have a page displaying the data table with filters applied, with the option to download the data as a report. A sketch of the proposed dashboard design is shown below.

(Spatiotemporal Analysis paragraph)

## Data Science Techniques

We will be using R Shiny to build the interactive dashboard, and we will be utilizing the leaflet and ggplot packages to assist in creating visually appealing graphs and maps. For data wrangling we will use the tidyverse set of packages to clean the data in preparation for plotting.

## Temporal Modeling

We will use a Bayesian temporal smoothing model for the inferential analysis of chronic disease incidence and prevalence over time. A Bayesian approach was selected in order to integrate prior knowledge about the

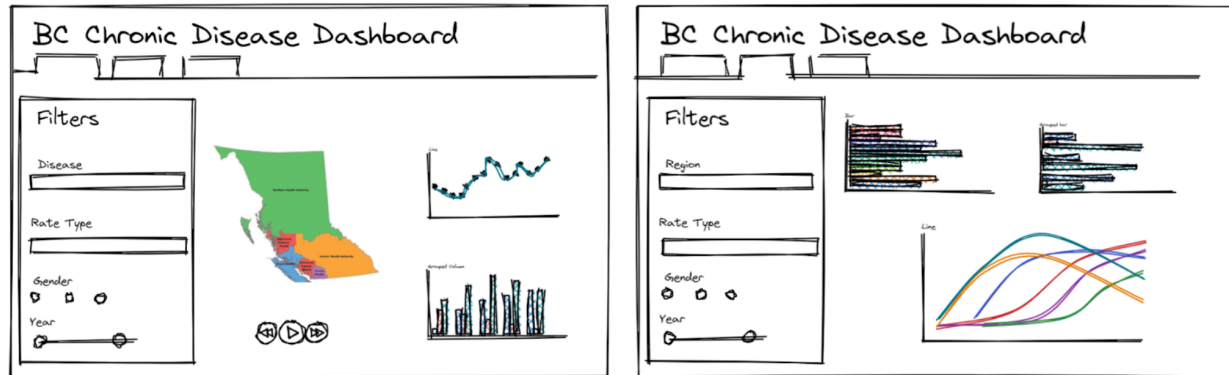


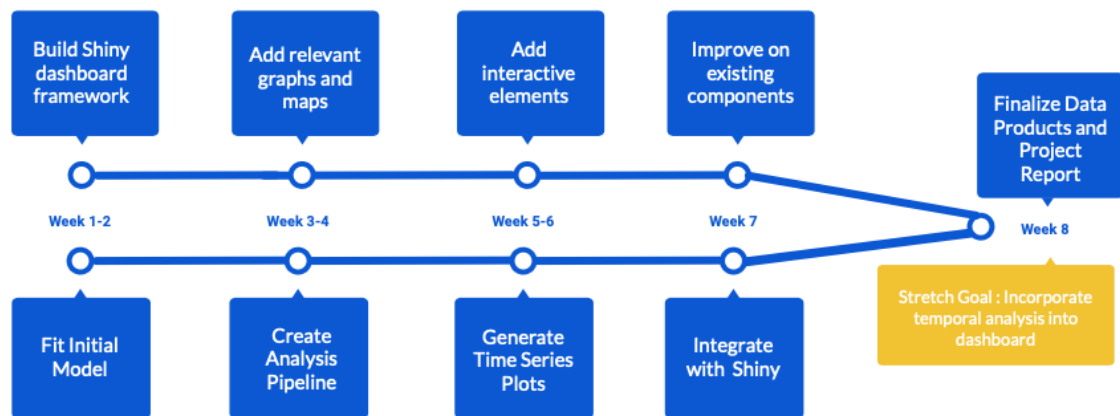
Figure 1: Proposed Dashboard Design

disease rates and to generate 95% credible intervals from the posterior distribution. Temporal smoothing is appropriate because we are assuming that the disease incidence and prevalence within a given region will be correlated over time. We propose to use the R-INLA package for Bayesian analysis for computational efficiency compared to traditional MCMC approaches.

For each disease, we will compare 3 temporal models for each of the 3 disease rates: an autoregressive model with order 1 (AR(1)), a random walk model with order 1 (RW(1)), and a random walk with order 2 (RW(2)). To select our final model, we will use the Deviance Information Criterion (DIC) and the Widely Applicable Information Criterion (WAIC), where a smaller DIC/WAIC indicates a better fit. After the best model is determined for that disease and metric, we will fit that model on all CHSA regions and generate plots with both observed and smoothened disease incidence/prevalence rates for the Shiny app.

## Timeline

### Dashboard



### Temporal Analysis

Figure 2: Timeline of Capstone project milestones

## Conclusion