Bayesian Data Analysis Project

```
knitr::opts_chunk$set(warning = FALSE, message = FALSE, error = TRUE)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x purrr::%||%()   masks base::%||%()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
theme_set(theme_bw())
require(extraDistr) #need for rdunif
```

```
## Loading required package: extraDistr
##
## Attaching package: 'extraDistr'
##
## The following object is masked from 'package:purrr':
##
##     rdunif
```

```
library(dplyr)
```

Introduction

In our world today, mental health has become more crucial and suicide remains a significant public health concern worldwide, with rates varying across regions due to complex social, economic, and cultural factors. Therefore, understanding the geographic distribution of suicide rates is important for the development of targeted mental health policies and prevention strategies. Although where we live in the world largely affects our mental and physical lifestyles, there has not been many studies done on the geographical characteristics of suicide rates and mental well-being.

In this study, we apply a Bayesian hierarchical model with a conditionally autoregressive (CAR) prior to investigate spatial patterns in suicide rates across countries, stratified by sex groups. We will be using a Poisson likelihood to obtain a prior distribution. After which, inferential analysis will be done using Monte Carlo Markov Chain (MCMC) done in Stan.

This research specifically focuses on two recent years,2019 and 2021, to investigate any observable changes in suicide patterns potentially influenced by global events like the COVID-19 pandemic. Our main research question is: Are there identifiable spatial clusters of high suicide rates, and do these clusters differ by sex?

This approach allows us to identify high-risk regions, quantify uncertainty, and better understand how neighboring countries may influence each other's suicide rates, providing valuable insights for data-informed mental health interventions.

Literature Review

Dataset and Data Cleaning

Dataset Name: Crude Suicide Rate (Per 100,000 Population) Source: https://www.who.int/data/gho/data/themes/mental-health/suicide-rates Description: The raw dataset has notable features like country, age group, sex, and suicide rate (per 100,000 people) that can be extracted.

Location: Country name Period: Year Dim1: Sex ("Female", "Both sexes", "Male) FactValueNumeric: Number of suicide deaths in a year, divided by the population and multiplied by 100 000 (as indicated in the original data source) FactValueNumericLow: Low estimate FactValueNumericHigh: High estimate

Note: The FactValueNumeric data are estimates of the number of suicides. "The estimates are derived from the WHO Global Health Estimates (GHE)" [data source]. However, some countries may not have an accurate way of recording the exact number of deaths, potentially leading to inaccurate estimations. Hence there is a high and low in the death rates. "For countries without high-quality death registration data, cause of death estimates are calculated using other data, including household surveys with verbal autopsy, sample or sentinel registration systems, special studies" [data source].

```
data_raw = read.csv("suicide_rate_raw.csv", header = TRUE)
head(data_raw)
```

```
##   IndicatorCode                                    Indicator ValueType
## 1    SDGSUICIDE Crude suicide rates (per 100 000 population)      text
## 2    SDGSUICIDE Crude suicide rates (per 100 000 population)      text
## 3    SDGSUICIDE Crude suicide rates (per 100 000 population)      text
## 4    SDGSUICIDE Crude suicide rates (per 100 000 population)      text
## 5    SDGSUICIDE Crude suicide rates (per 100 000 population)      text
## 6    SDGSUICIDE Crude suicide rates (per 100 000 population)      text
##   ParentLocationCode        ParentLocation Location.type SpatialDimValueCode
## 1                AMR              Americas       Country                 VCT
## 2                EMR Eastern Mediterranean       Country                 OMN
## 3                EMR Eastern Mediterranean       Country                 PSE
## 4                EMR Eastern Mediterranean       Country                 JOR
## 5                EMR Eastern Mediterranean       Country                 KWT
## 6                EMR Eastern Mediterranean       Country                 SYR
##                                                Location Period.type Period
## 1                       Saint Vincent and the Grenadines        Year   2021
## 2                                                  Oman        Year   2021
## 3 occupied Palestinian territory, including east Jerusalem        Year   2021
## 4                                                Jordan        Year   2021
## 5                                                Kuwait        Year   2021
## 6                                  Syrian Arab Republic        Year   2021
##   IsLatestYear Dim1.type   Dim1 Dim1ValueCode Dim2.type                  Dim2
## 1         true       Sex Female      SEX_FMLE Age Group All age groups (total)
## 2         true       Sex Female      SEX_FMLE Age Group All age groups (total)
## 3         true       Sex Female      SEX_FMLE Age Group All age groups (total)
## 4         true       Sex Female      SEX_FMLE Age Group All age groups (total)
```

```
## 5         true       Sex Female      SEX_FMLE Age Group All age groups (total)
## 6         true       Sex Female      SEX_FMLE Age Group All age groups (total)
##       Dim2ValueCode Dim3.type Dim3 Dim3ValueCode DataSourceDimValueCode
## 1 AGEGROUP_YEARSALL        NA   NA            NA                     NA
## 2 AGEGROUP_YEARSALL        NA   NA            NA                     NA
## 3 AGEGROUP_YEARSALL        NA   NA            NA                     NA
## 4 AGEGROUP_YEARSALL        NA   NA            NA                     NA
## 5 AGEGROUP_YEARSALL        NA   NA            NA                     NA
## 6 AGEGROUP_YEARSALL        NA   NA            NA                     NA
##   DataSource FactValueNumericPrefix FactValueNumeric FactValueUoM
## 1         NA                     NA             0.00           NA
## 2         NA                     NA             0.21           NA
## 3         NA                     NA             0.23           NA
## 4         NA                     NA             0.29           NA
## 5         NA                     NA             0.34           NA
## 6         NA                     NA             0.26           NA
##   FactValueNumericLowPrefix FactValueNumericLow FactValueNumericHighPrefix
## 1                        NA                0.00                         NA
## 2                        NA                0.12                         NA
## 3                        NA                0.14                         NA
## 4                        NA                0.18                         NA
## 5                        NA                0.25                         NA
## 6                        NA                0.16                         NA
##   FactValueNumericHigh         Value FactValueTranslationID FactComments
## 1                0.00 0.0 [0.0-0.0]                     NA           NA
## 2                0.34 0.2 [0.1-0.3]                     NA           NA
## 3                0.36 0.2 [0.1-0.4]                     NA           NA
## 4                0.43 0.3 [0.2-0.4]                     NA           NA
## 5                0.39 0.3 [0.2-0.4]                     NA           NA
## 6                0.42 0.3 [0.2-0.4]                     NA           NA
##   Language             DateModified
## 1       EN 2025-01-10T08:00:00.000Z
## 2       EN 2025-01-10T08:00:00.000Z
## 3       EN 2025-01-10T08:00:00.000Z
## 4       EN 2025-01-10T08:00:00.000Z
## 5       EN 2025-01-10T08:00:00.000Z
## 6       EN 2025-01-10T08:00:00.000Z
```

```
data = as.data.frame(data_raw |> select(Location, Period, Dim1, FactValueNumeric, FactValueNumericLow, F
head(data)
```

```
##                                                Location Period   Dim1
## 1                        Saint Vincent and the Grenadines   2021 Female
## 2                                                   Oman   2021 Female
## 3 occupied Palestinian territory, including east Jerusalem   2021 Female
## 4                                                 Jordan   2021 Female
## 5                                                 Kuwait   2021 Female
## 6                                   Syrian Arab Republic   2021 Female
##   FactValueNumeric FactValueNumericLow FactValueNumericHigh
## 1             0.00                0.00                 0.00
## 2             0.21                0.12                 0.34
## 3             0.23                0.14                 0.36
## 4             0.29                0.18                 0.43
## 5             0.34                0.25                 0.39
```

```
## 6             0.26                0.16                0.42
```

```
unique(data$Period)
```

```
##  [1] 2021 2020 2019 2018 2017 2016 2015 2014 2013 2012 2011 2010 2009 2008 2007
## [16] 2006 2005 2004 2003 2002 2001 2000
```

filter data to only include 2019 and 2021

```
suicide_2019 <- filter(data, Period == 2019)
suicide_2021 <- filter(data, Period == 2021)
```

Data Analysis

As we have obtained the cleaned data for suicide rates in 2019 and 2021, we can now declare a prior model from information obtained historically.