Exploration

Table of contents

## Load Libraries

library(tidyverse)  
library(lubridate)  
library(fs)  
library(tsibble)  
library(fpp3)

## Load and Clean Data

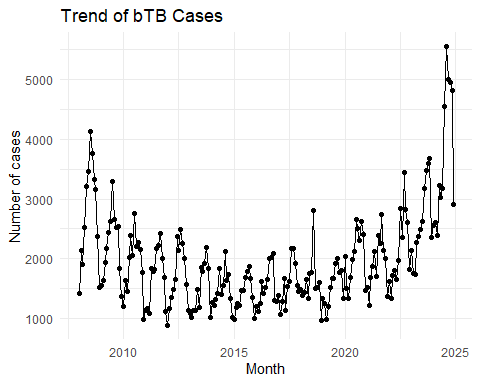
all\_cases\_collapsed\_file <- dir\_info("data", type = "file", regexp = "(?i)all\_cases\_collapsed.\*\\.rds$") %>%  
 subset(subset = modification\_time == max(modification\_time), select = path, drop = TRUE)  
  
all\_cases\_collapsed <- readRDS(all\_cases\_collapsed\_file) %>%  
 filter(year > 2007, year != 2025) %>%  
 as.data.frame()

##Aggregate by Month

cases\_per\_month <- all\_cases\_collapsed %>%  
 group\_by(year\_month) %>%  
 summarise(num\_cases = n(), .groups = "drop") %>%  
 mutate(  
 year\_month = as.Date(paste0(year\_month, "-01")),  
 year = year(year\_month),  
 month = month(year\_month)  
 )

##General Trend of bTB Cases

ggplot(cases\_per\_month, aes(x = year\_month, y = num\_cases)) +  
 geom\_line() +  
 geom\_point() +  
 labs(title = "Trend of bTB Cases", x = "Month", y = "Number of cases") +  
 theme\_minimal()



##Average Cases per Month

avg\_per\_month <- cases\_per\_month %>%  
 group\_by(month) %>%  
 summarise(avg\_cases = mean(num\_cases), .groups = "drop")

##Monthly Seasonality Plot

ggplot(cases\_per\_month, aes(x = year, y = num\_cases, group = 1)) +  
 geom\_line() +  
 geom\_point(size = 1) +  
 geom\_hline(  
 data = avg\_per\_month,  
 aes(yintercept = avg\_cases),  
 color = "blue",  
 linetype = "solid"  
 ) +  
 facet\_grid(. ~ month, scales = "free\_x", space = "free") +  
 labs(  
 title = glue::glue("Seasonality of bTB by month (2008-{max(cases\_per\_month$year)})"),  
 x = "Year",  
 y = "Number of cases"  
 ) +  
 theme\_minimal() +  
 theme(  
 strip.background = element\_rect(fill = "grey90", color = NA),  
 strip.text = element\_text(size = 12, face = "bold"),  
 axis.text.x = element\_text(angle = 90, hjust = 1, face = "bold")  
 )

