STAT 447C Project v0

Jerry Gai, Sunny Lau

Due April 19

1 Part 1: Problem Formulation (15%)

The report clearly describes:

- a real-world inference task/problem,
- succinct but sufficient context (e.g. biological terminology) needed to understand the problem,
- the key modelling/methodological/challenge, clearly associating it with one of the items under "menu of project themes" above.

2 Part 2: a literature review: (10%)

- relevant literature is cited and properly summarized.

3 Part 3: Data analysis (40%)

- A Bayesian model is precisely described (e.g. using the .. ~ .. notation)
- Implementation code in the appendix (e.g. using Stan)
- Prior choice is motivated. If appropriate, several choices are compared or sensitivity analysis is performed.
- Critical evaluation of the posterior approximation. An appropriate combination of diagnostics, synthetic datasets and other validation strategies.

3.1 Plotting data

```
eq_data <- read.csv("earthquake_california/db_data.csv",
                    stringsAsFactors = FALSE, strip.white = TRUE)

eq_data$DateTime <- as.POSIXct(paste(eq_data$Date, eq_data$Time),
                               format = "%Y/%m/%d %H:%M:%OS", tz = "UTC")

# create a weekly bin
eq_data$Week <- floor_date(eq_data$DateTime, unit = "week")

# aggregate the event counts by week.
# can adjust threshold of `Mag`, 3.0 seems too many earthquake
```
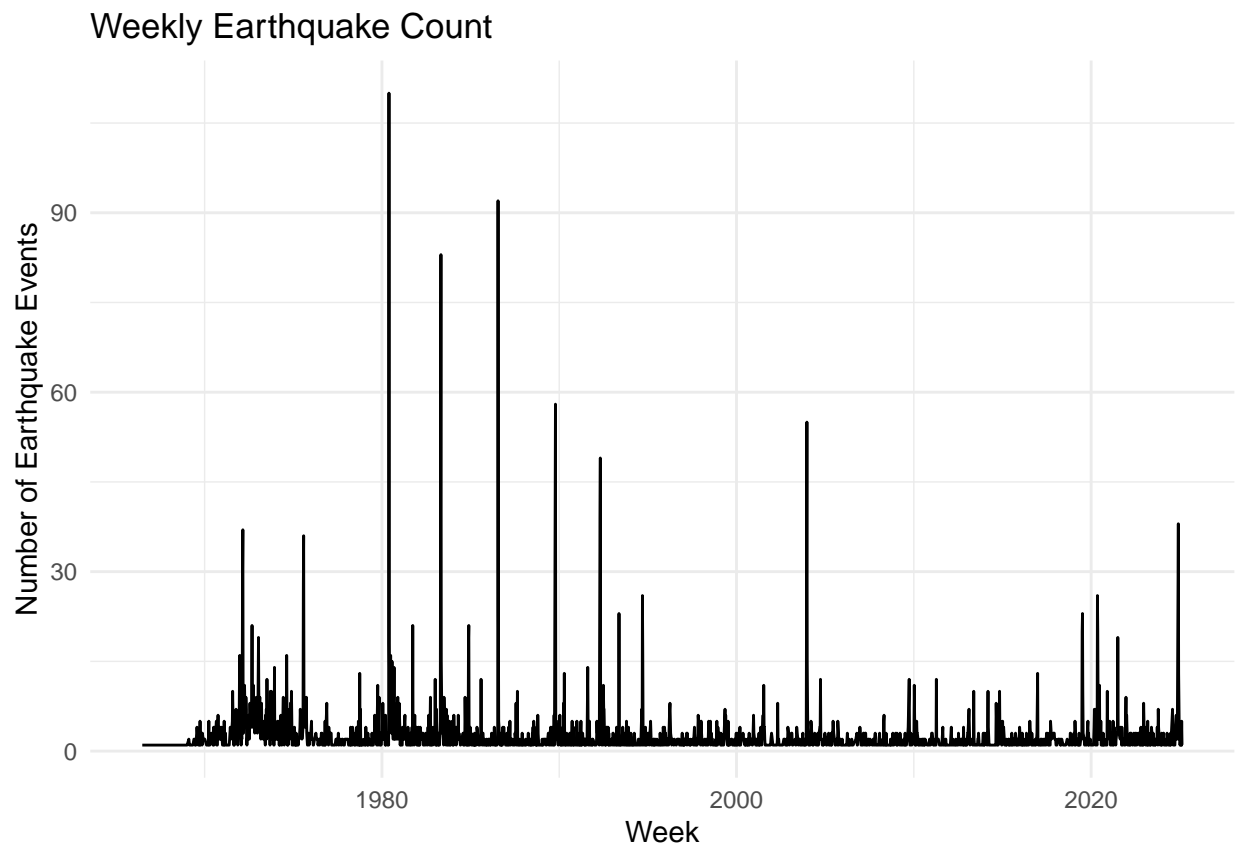
```
weekly_counts <- eq_data %>%
  group_by(Week) %>%
  filter(Mag > 3.5) %>%
  summarise(Count = n())

ggplot(weekly_counts, aes(x = Week, y = Count)) +
  geom_line() +
  labs(title = "Weekly Earthquake Count",
       x = "Week",
       y = "Number of Earthquake Events") +
  theme_minimal()
```


4 Part 4: Project theme: methodological/theoretical aspect of the project (20%)

- Is the approach sound?
- Creative?

5 Part 5: Discussion (5%)

- Does the report describe key limitations?