

Week 10: Temporal data

26/03/23

Child mortality in Sri Lanka

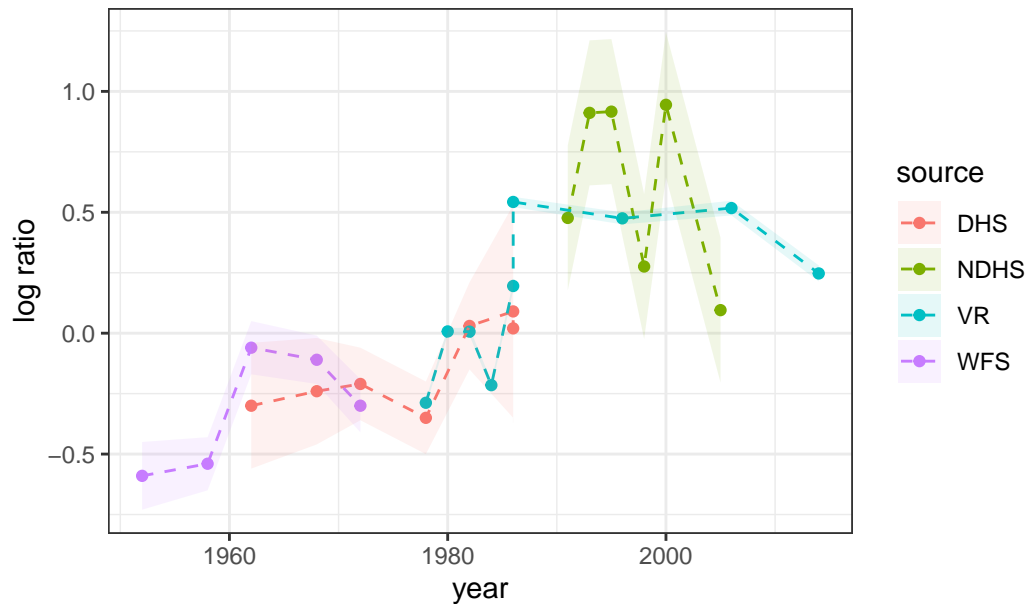
In this lab you will be fitting a couple of different models to the data about child mortality in Sri Lanka, which was used in the lecture. Here's the data and the plot from the lecture:

```
knitr::opts_chunk$set(results = 'hide', message = F, warning = F)

library(tidyverse)
library(here)
library(rstan)
library(tidybayes)

lka <- read_csv(here("Labs/Lab7/lka.csv"))
ggplot(lka, aes(year, logit_ratio)) +
  geom_point(aes( color = source)) +
  geom_line(aes( color = source), lty = 2) +
  geom_ribbon(aes(ymin = logit_ratio - se,
                 ymax = logit_ratio + se,
                 fill = source), alpha = 0.1) +
  theme_bw()+
  labs(title = "Ratio of neonatal to other child mortality (logged), Sri Lanka", y = "log
```

Ratio of neonatal to other child mortality (logged), Sri Lanka



Fitting a linear model

Let's firstly fit a linear model in time to these data. Here's the code to do this:

```
observed_years <- lka$year
years <- min(observed_years):max(observed_years)
nyears <- length(years)

stan_data <- list(y = lka$logit_ratio, year_i = observed_years - years[1]+1,
                  T = nyears, years = years, N = length(observed_years),
                  mid_year = mean(years), se = lka$se)

mod <- stan(data = stan_data,
            file = here("Labs/Lab7/lka_linear_me.stan"))
```

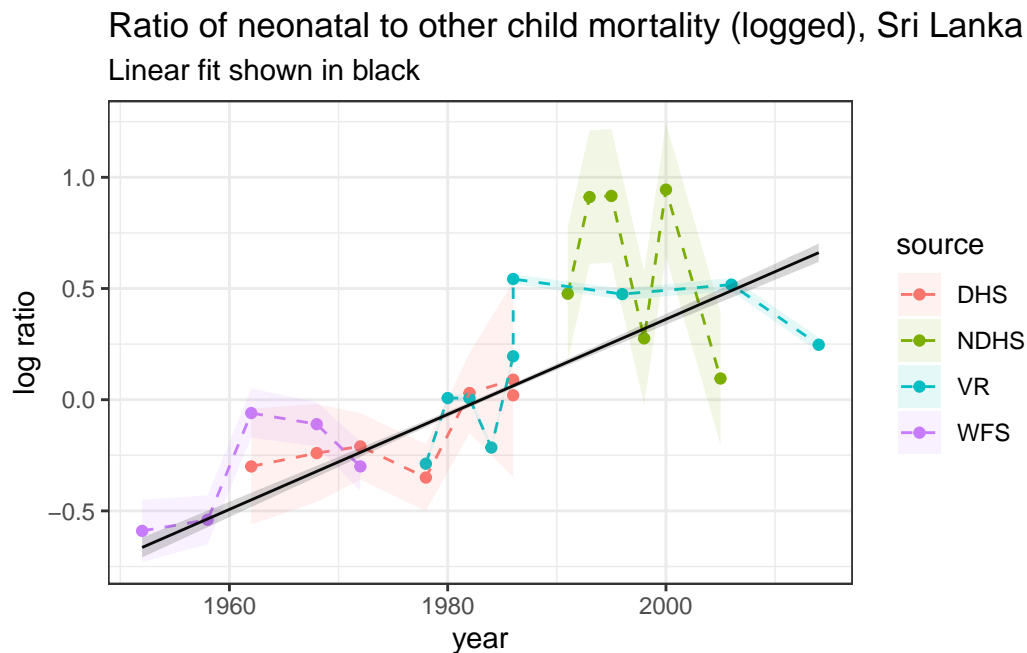
Extract the results:

```
res <- mod %>%
  gather_draws(mu[t]) %>%
  median_qi() %>%
  mutate(year = years[t])
```

Plot the results:

```
ggplot(lka, aes(year, logit_ratio)) +
  geom_point(aes( color = source)) +
  geom_line(aes( color = source), lty = 2) +
  geom_ribbon(aes(ymin = logit_ratio - se,
                 ymax = logit_ratio + se,
                 fill = source), alpha = 0.1) +

  theme_bw()+
  geom_line(data = res, aes(year, .value)) +
  geom_ribbon(data = res, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2)+
  theme_bw()+
  labs(title = "Ratio of neonatal to other child mortality (logged), Sri Lanka",
       y = "log ratio", subtitle = "Linear fit shown in black")
```



Question 1

Project the linear model above out to 2023 by adding a **generated quantities** block in Stan (do the projections based on the expected value μ). Plot the resulting projections on a graph similar to that above.

```
stan_data <- list(y = lka$logit_ratio, year_i = observed_years - years[1]+1,
                 T = nyears, years = years, N = length(observed_years),
                 mid_year = mean(years), se = lka$se,P=9)
```

```
mod2 <- stan(data = stan_data,
             file = here("Labs/Lab7/lka2.stan"))
#mod2
```

```
res_linear=mod2 %>%
  gather_draws(mu[t]) %>%
  median_qi() %>%
  mutate(year=years[t])
```

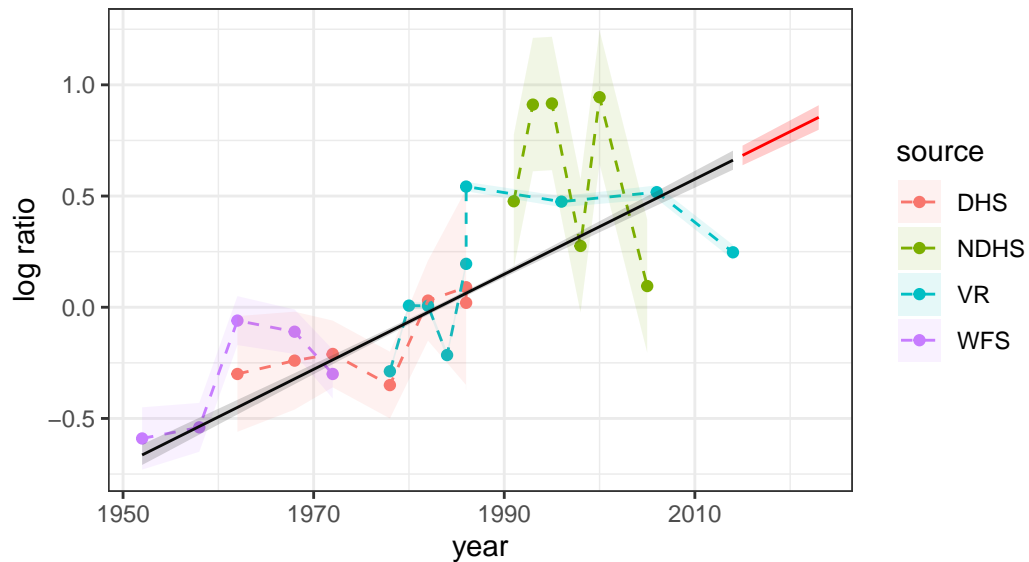
```
res_linear_p=mod2 %>%
  gather_draws(mu_p[p]) %>%
  median_qi() %>%
  mutate(year=years[nyears]+p)
```

```
ggplot(lka, aes(year, logit_ratio)) +
  geom_point(aes( color = source)) +
  geom_line(aes( color = source), lty = 2) +
  geom_ribbon(aes(ymin = logit_ratio - se,
                 ymax = logit_ratio + se,
                 fill = source), alpha = 0.1) +

  theme_bw()+
  geom_line(data = res_linear, aes(year, .value)) +
  geom_ribbon(data = res_linear, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.1) +
  geom_line(data = res_linear_p, aes(year, .value),col="red") +
  geom_ribbon(data = res_linear_p, aes(y = .value, ymin = .lower, ymax = .upper),fill="red") +

  theme_bw()+
  labs(title = "Ratio of neonatal to other child mortality (logged), Sri Lanka",
       y = "log ratio", subtitle = "Linear fit shown in black, projections in red")
```

Ratio of neonatal to other child mortality (logged), Sri Lanka
Linear fit shown in black, projections in red



Random walks

Question 2

Code up and estimate a first order random walk model to fit to the Sri Lankan data, taking into account measurement error, and project out to 2023.

```
mod3 <- stan(data = stan_data,
             file = here("Labs/Lab7/lka3.stan"))
#mod3
```

```
res=mod3 %>%
  gather_draws(mu[t]) %>%
  median_qi() %>%
  mutate(year=years[t])

res_p=mod3 %>%
  gather_draws(mu_p[p]) %>%
  median_qi() %>%
  mutate(year=years[nyears]+p)
```

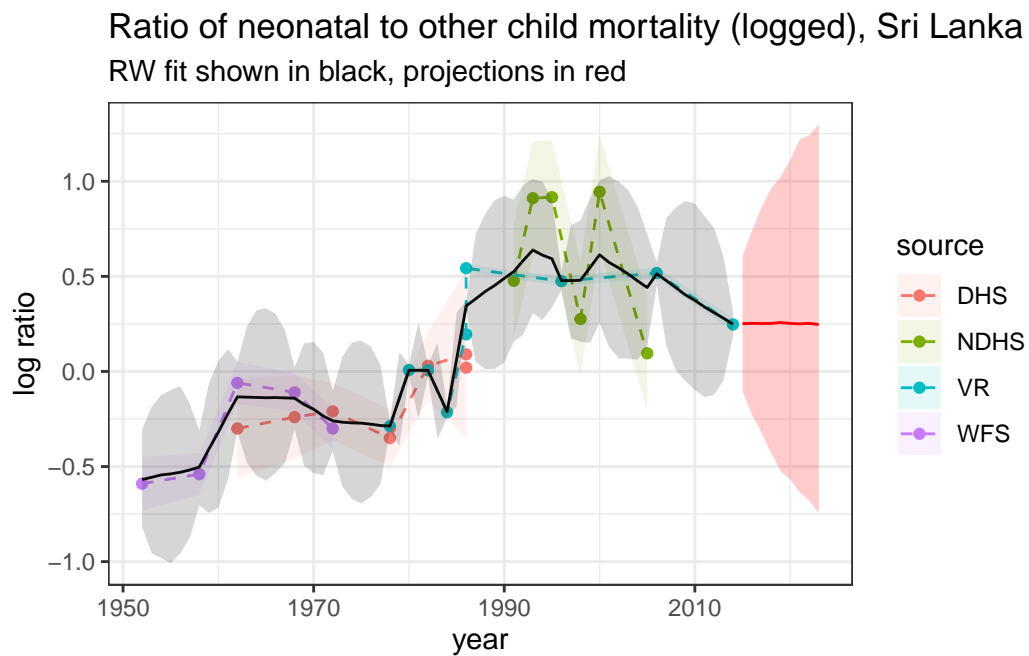
```

ggplot(lka, aes(year, logit_ratio)) +
  geom_point(aes( color = source)) +
  geom_line(aes( color = source), lty = 2) +
  geom_ribbon(aes(ymin = logit_ratio - se,
                 ymax = logit_ratio + se,
                 fill = source), alpha = 0.1) +

  theme_bw()+
  geom_line(data = res, aes(year, .value)) +
  geom_ribbon(data = res, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2)+
  geom_line(data = res_p, aes(year, .value), col="red") +
  geom_ribbon(data = res_p, aes(y = .value, ymin = .lower, ymax = .upper), fill="red", alpha = 0.2)+

  theme_bw()+
  labs(title = "Ratio of neonatal to other child mortality (logged), Sri Lanka",
       y = "log ratio", subtitle = "RW fit shown in black, projections in red")

```



Question 3

Now alter your model above to estimate and project a second-order random walk model (RW2).

Question 4

Run the first order and second order random walk models, including projections out to 2023. Compare these estimates with the linear fit by plotting everything on the same graph.

```
mod4 <- stan(data = stan_data,
             file = here("Labs/Lab7/lka4.stan"))
#mod4

# Plot projections
res2=mod4 %>%
  gather_draws(mu[t]) %>%
  median_qi() %>%
  mutate(year=years[t])

res2_p=mod4 %>%
  gather_draws(mu_p[p]) %>%
  median_qi() %>%
  mutate(year=years[nyears]+p)

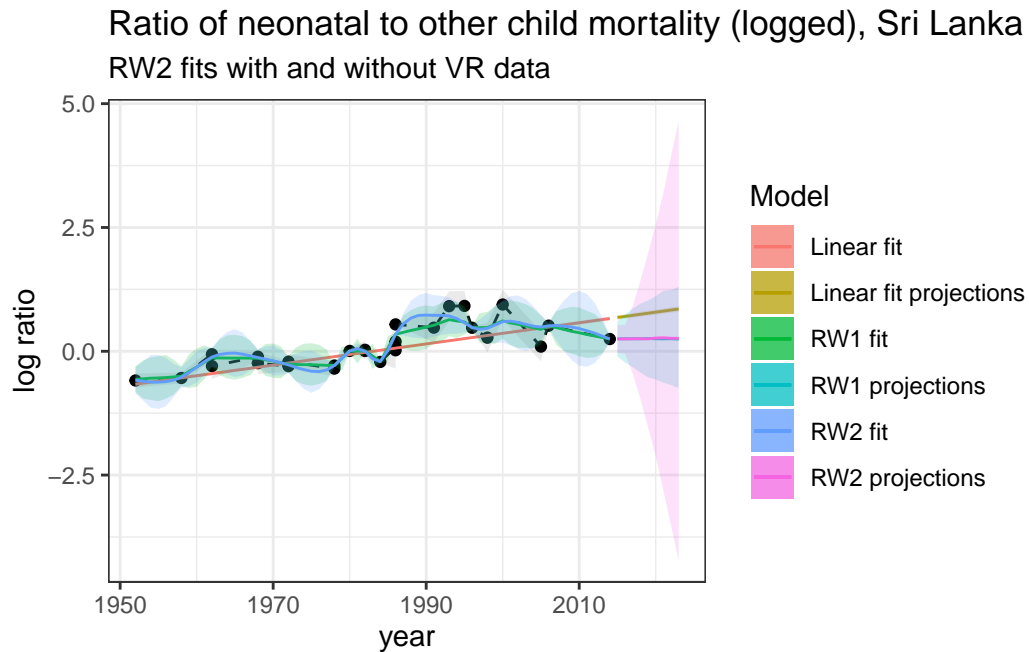
ggplot(lka, aes(year, logit_ratio)) +
  geom_point() +
  geom_line(lty = 2) +
  geom_ribbon(aes(ymin = logit_ratio - se,
                ymax = logit_ratio + se), alpha = 0.1) +
  geom_line(data = res_linear, aes(year, .value, col = "Linear fit")) +
  geom_ribbon(data = res_linear, aes(y = .value, ymin = .lower, ymax = .upper, fill = "Linear fit")) +
  geom_line(data = res_linear_p, aes(year, .value,col="Linear fit projections")) +
  geom_ribbon(data = res_linear_p, aes(y = .value, ymin = .lower, ymax = .upper,fill="Linear fit projections")) +
  geom_line(data = res, aes(year, .value, col = "RW1 fit")) +
  geom_ribbon(data = res, aes(y = .value, ymin = .lower, ymax = .upper, fill = "RW1 fit"),
  geom_line(data = res_p, aes(year, .value,col="RW1 projections")) +
  geom_ribbon(data = res_p, aes(y = .value, ymin = .lower, ymax = .upper,fill="RW1 projections")) +

  geom_line(data = res2, aes(year, .value, col = "RW2 fit")) +
  geom_ribbon(data = res2, aes(y = .value, ymin = .lower, ymax = .upper, fill = "RW2 fit")) +

  geom_line(data = res2_p, aes(year, .value,col="RW2 projections")) +
  geom_ribbon(data = res2_p, aes(y = .value, ymin = .lower, ymax = .upper,fill="RW2 projections")) +

  theme_bw() +
  labs(title = "Ratio of neonatal to other child mortality (logged), Sri Lanka",
```

```
y = "log_ratio", subtitle = "RW2 fits with and without VR data", col = "Model", fi
```



Question 5

Rerun the RW2 model excluding the VR data. Briefly comment on the differences between the two data situations.

```
lka_no_VR <- lka %>% filter(source != "VR")
observed_years <- lka_no_VR$year
years <- min(observed_years):max(observed_years)
nyears <- length(years)

stan_data <- list(y = lka_no_VR$logit_ratio, year_i = observed_years - years[1]+1,
                  T = nyears, years = years, N = length(observed_years),
                  mid_year = mean(years), se = lka_no_VR$se, P = 18)
mod5 <- stan(data = stan_data,
             file = here("Labs/Lab7/lka4.stan"))

#mod5

# Plot projections
res3=mod5 %>%
```



```

gather_draws(mu[t]) %>%
median_qi() %>%
mutate(year=years[t])

res3_p=mod5 %>%
gather_draws(mu_p[p]) %>%
median_qi() %>%
mutate(year=years[nyears]+p)

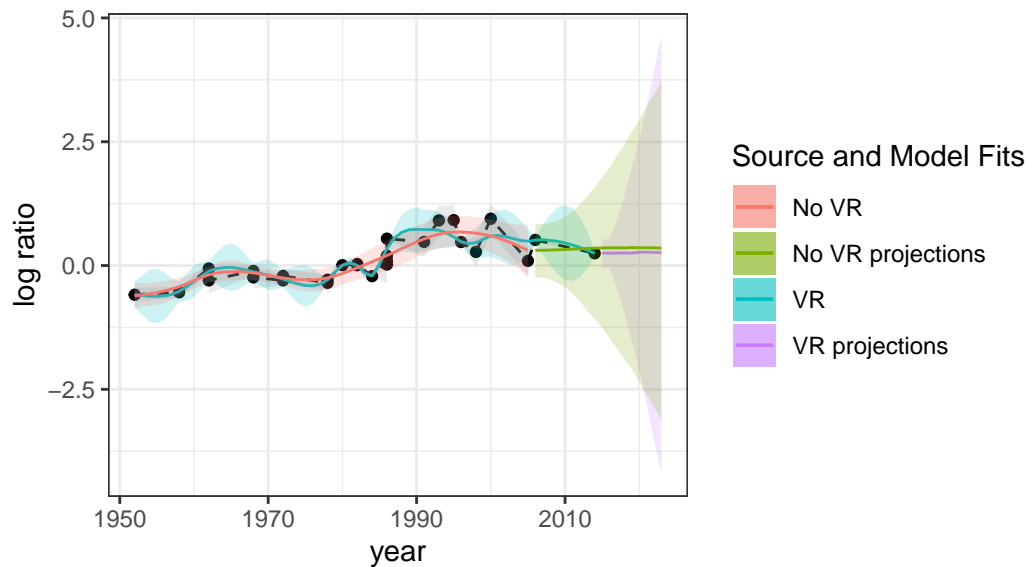
ggplot(lka, aes(year, logit_ratio)) +
  geom_point() +
  geom_line(lty = 2) +
  geom_ribbon(aes(ymin = logit_ratio - se,
                  ymax = logit_ratio + se), alpha = 0.1) +
  geom_line(data = res2, aes(year, .value, col = "VR")) +
  geom_ribbon(data = res2, aes(y = .value, ymin = .lower, ymax = .upper, fill = "VR"), alp
  geom_line(data = res2_p, aes(year, .value,col="VR projections")) +
  geom_ribbon(data = res2_p, aes(y = .value, ymin = .lower, ymax = .upper,fill="VR project

  geom_line(data = res3, aes(year, .value, col = "No VR")) +
  geom_ribbon(data = res3, aes(y = .value, ymin = .lower, ymax = .upper, fill = "No VR"),

  geom_line(data = res3_p, aes(year, .value,col="No VR projections")) +
  geom_ribbon(data = res3_p, aes(y = .value, ymin = .lower, ymax = .upper,fill="No VR proje
  theme_bw()+
  labs(title = "Ratio of neonatal to other child mortality (logged), Sri Lanka",
        y = "log ratio", subtitle = "RW2 fits with and without VR data", col = "Source and

```

Ratio of neonatal to other child mortality (logged), Sri Lanka
RW2 fits with and without VR data



We note that by removing the VR data, the 95% credible intervals around the fitted values is smaller while still covering the true observed values. However, now the latest data point is from 2005 which means that we have to generate predictions for the years 2006 to 2014. Nevertheless, the model fitted without the VR data also produces a smaller credible interval for the latest years (2020 onwards).

Question 6

Briefly comment on which model you think is most appropriate, or an alternative model that would be more appropriate in this context.

From the results in Question 4, RW1 appears to be a better fit since the projections and fitted values are close to that of the RW2 while also having smaller credible intervals.