lab10

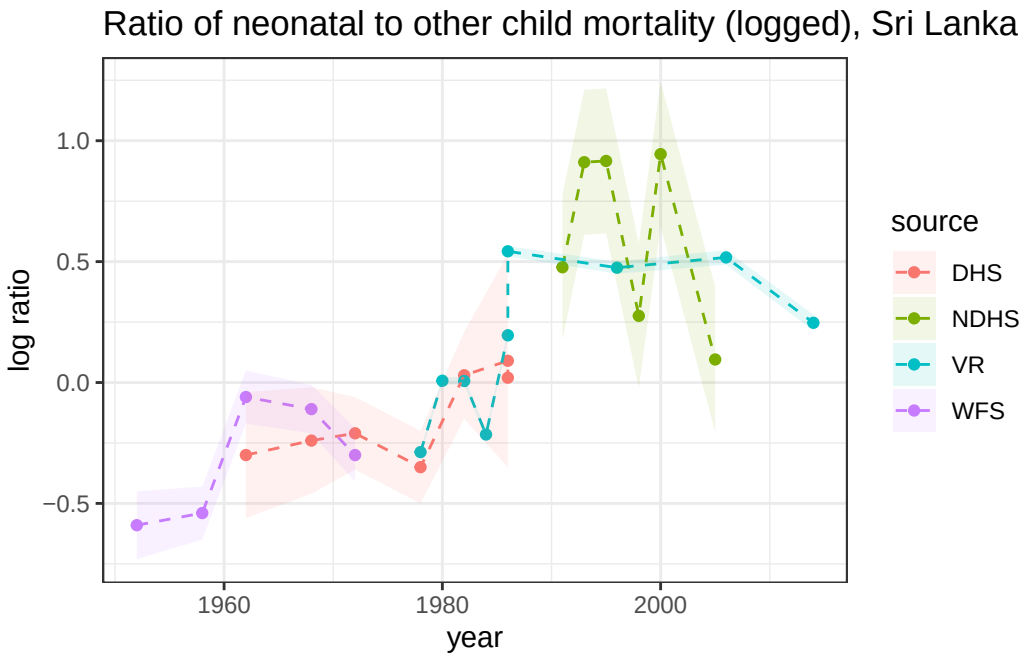
27/03/23

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:

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

```
lka <- read_csv(here("data/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"
```



Fitting a linear model

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

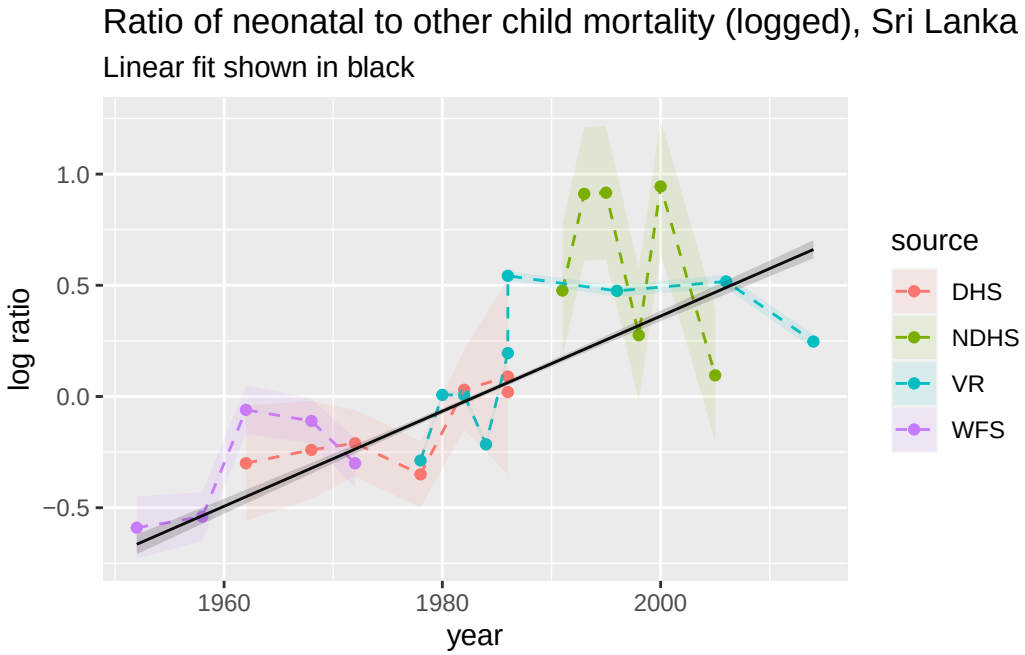
```
observed_years <- lka$year
years <- min(observed_years):max(observed_years)
nyears <- length(years)

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("code/models/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)+
  geom_line(data = res, aes(year, .value)) +
  geom_ribbon(data = res, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2)+labs
       y = "log ratio", subtitle = "Linear fit shown in black")
```



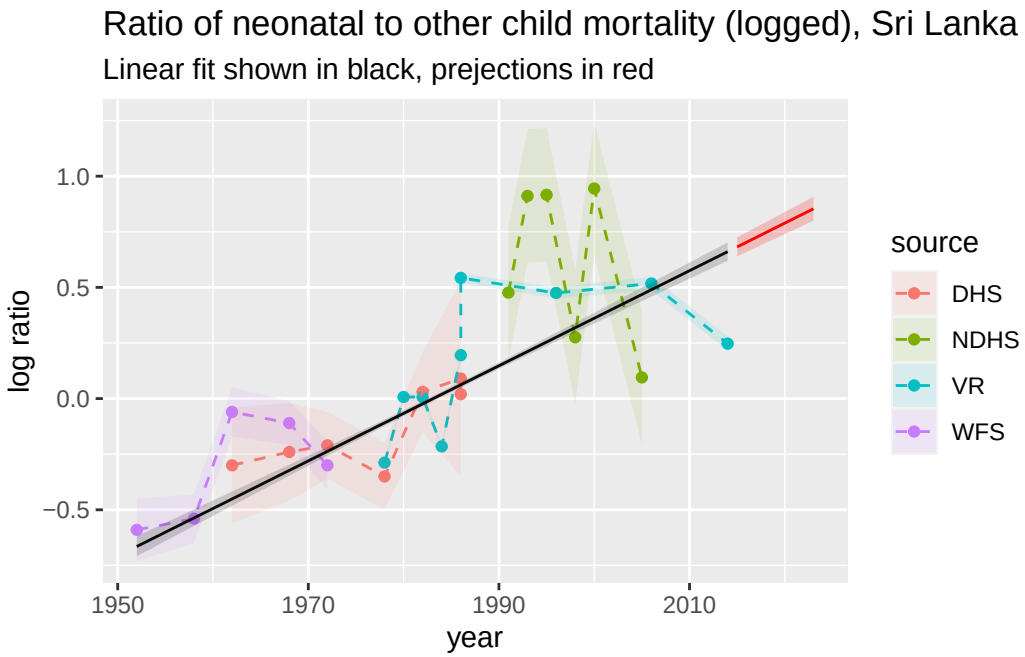
Question 1

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_p <- 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)
```

```
mod1 <- stan(data = stan_data_p,
             file = here("code/models/lab10_1.stan"))
```

```
res1 <- mod1 %>%
  gather_draws(mu[t]) %>%
  median_qi() %>%
  mutate(year = years[t])

res1_p <- mod1 %>%
  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) +
  geom_line(data = res1, aes(year, .value)) +
  geom_ribbon(data = res1, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2)+
geom_line(data = res1_p, aes(year, .value),col="red") +
  geom_ribbon(data = res1_p, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2,fi
       y = "log ratio", subtitle = "Linear fit shown in black, prejections in red")
```



Random walks

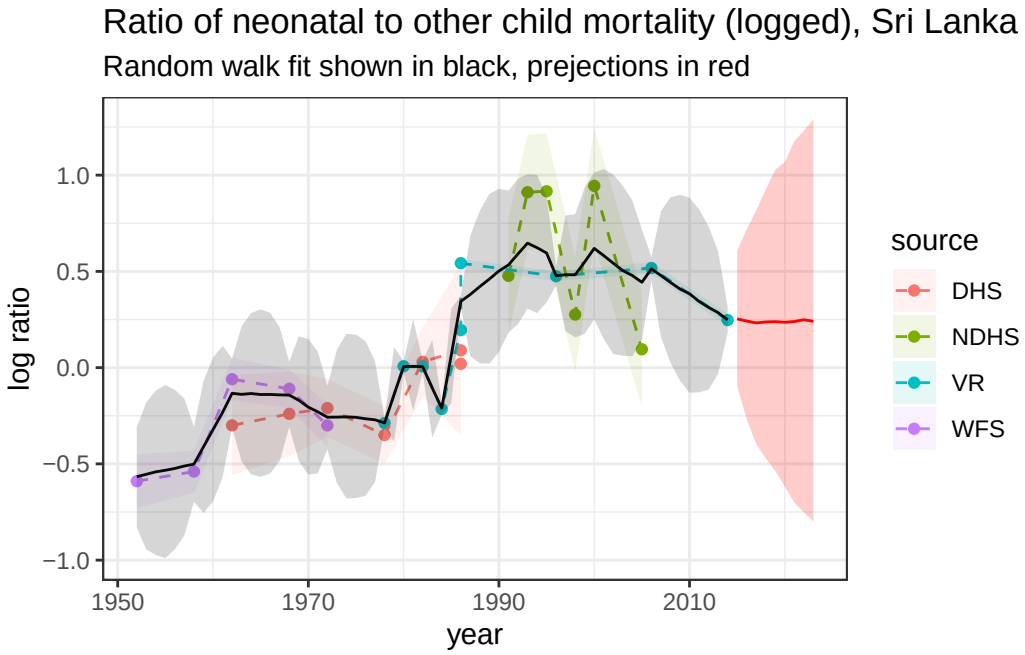
Question 2

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.

```
mod2 <- stan(data = stan_data_p,
             file = here("code/models/lab10_2.stan"))
```

```
res2 <- mod2 %>%
  gather_draws(mu[t]) %>%
  median_qi() %>%
  mutate(year = years[t])

res2_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 = res2, aes(year, .value)) +
  geom_ribbon(data = res2, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2)+
geom_line(data = res2_p, aes(year, .value),col="red") +
  geom_ribbon(data = res2_p, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2,fi
       y = "log ratio", subtitle = "Random walk fit shown in black, prejections in red")
```

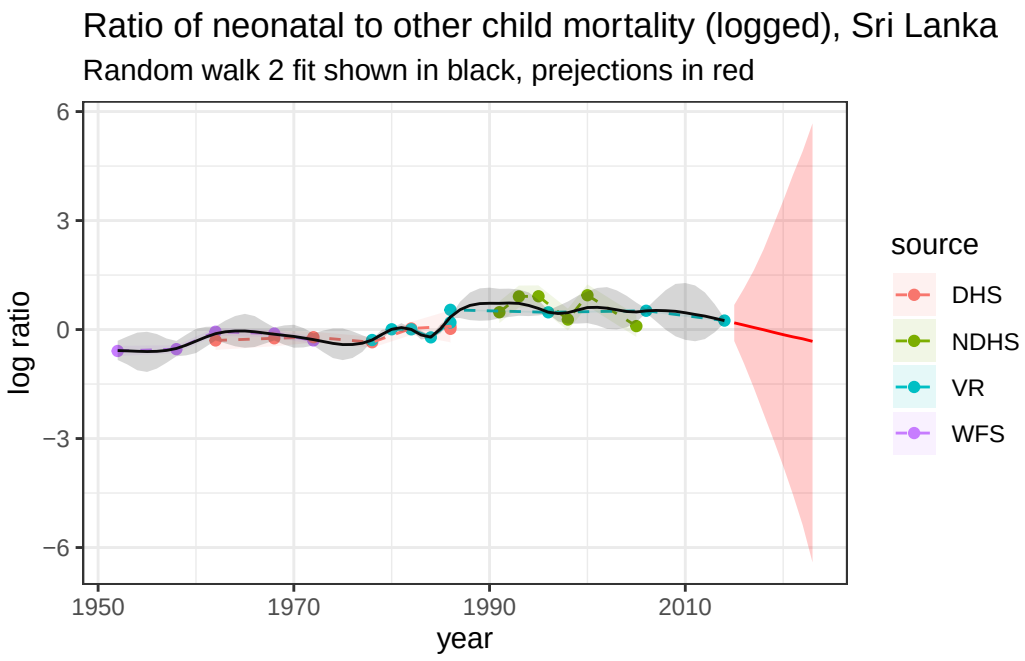


Question 3

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

```
mod3 <- stan(data = stan_data_p,
             file = here("code/models/lab10_3.stan"))
```

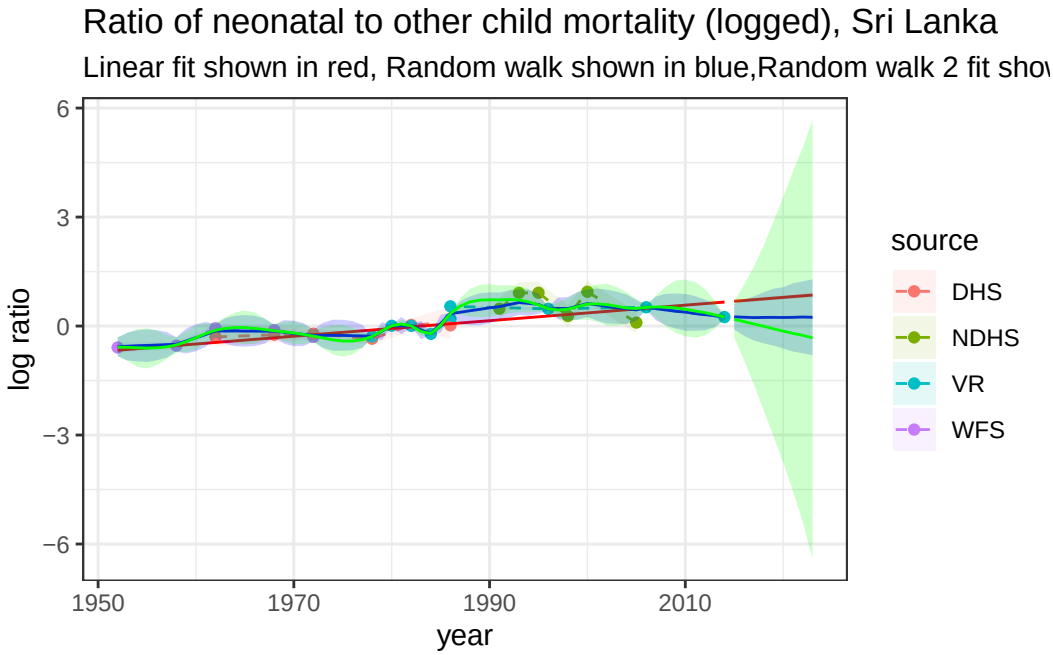
```r
res3 <- mod3 %>%
  gather_draws(mu[t]) %>%
  median_qi() %>%
  mutate(year = years[t])

res3_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 = res3, aes(year, .value)) +
  geom_ribbon(data = res3, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2)+
geom_line(data = res3_p, aes(year, .value),col="red") +
  geom_ribbon(data = res3_p, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2,fi
       y = "log ratio", subtitle = "Random walk 2 fit shown in black, prejections in red")
```



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.

```
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 = res1, aes(year, .value),col="red") +
  geom_ribbon(data = res1, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2,fill
geom_line(data = res1_p, aes(year, .value),col="red") +
  geom_ribbon(data = res1_p, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2,fi
  geom_line(data = res2, aes(year, .value),col="blue") +
  geom_ribbon(data = res2, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2,fill
geom_line(data = res2_p, aes(year, .value),col="blue") +
  geom_ribbon(data = res2_p, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2,fi
  geom_line(data = res3, aes(year, .value),col="green") +
  geom_ribbon(data = res3, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2,fill
geom_line(data = res3_p, aes(year, .value),col="green") +
  geom_ribbon(data = res3_p, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2,fi
       y = "log ratio", subtitle = "Linear fit shown in red, Random walk shown in blue,Ran
```



Question 5

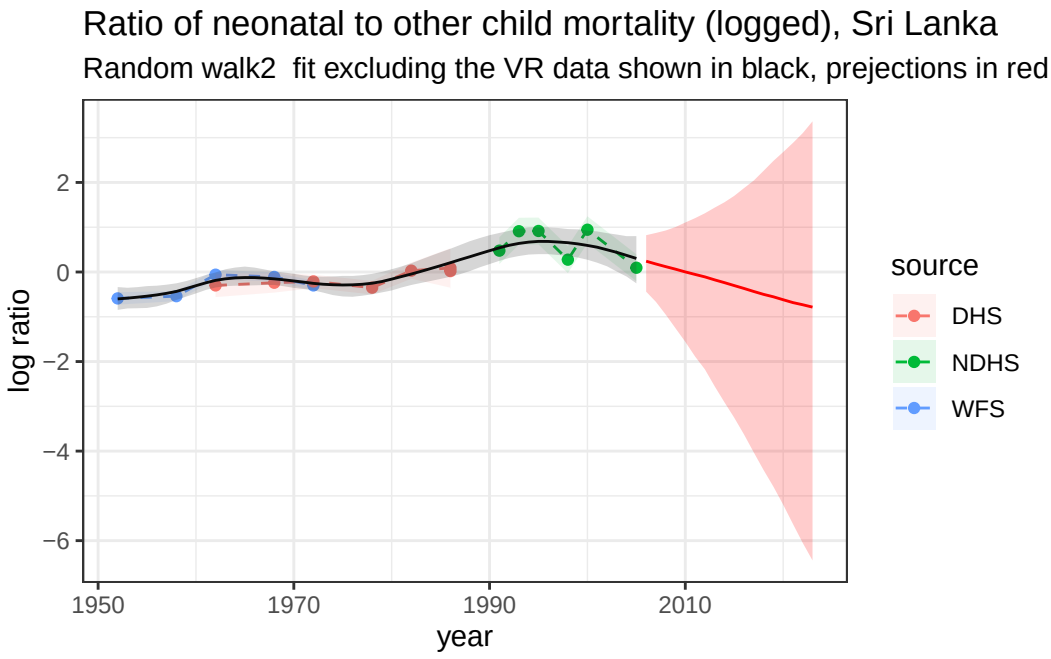
Rerun the RW2 model excluding the VR data. Briefly comment on the differences between the two data situations.

```
lka2 <- lka|>
  filter(source!="VR")
```

```
observed_years<- lka2$year
years <- min(observed_years):max(observed_years)
nyears <- length(years)

stan_data2 <- list(y = lka2$logit_ratio, year_i = observed_years - years[1]+1,
                   T = nyears, years = years, N = length(observed_years),
                   mid_year = mean(years), se = lka2$se,P=18)
```

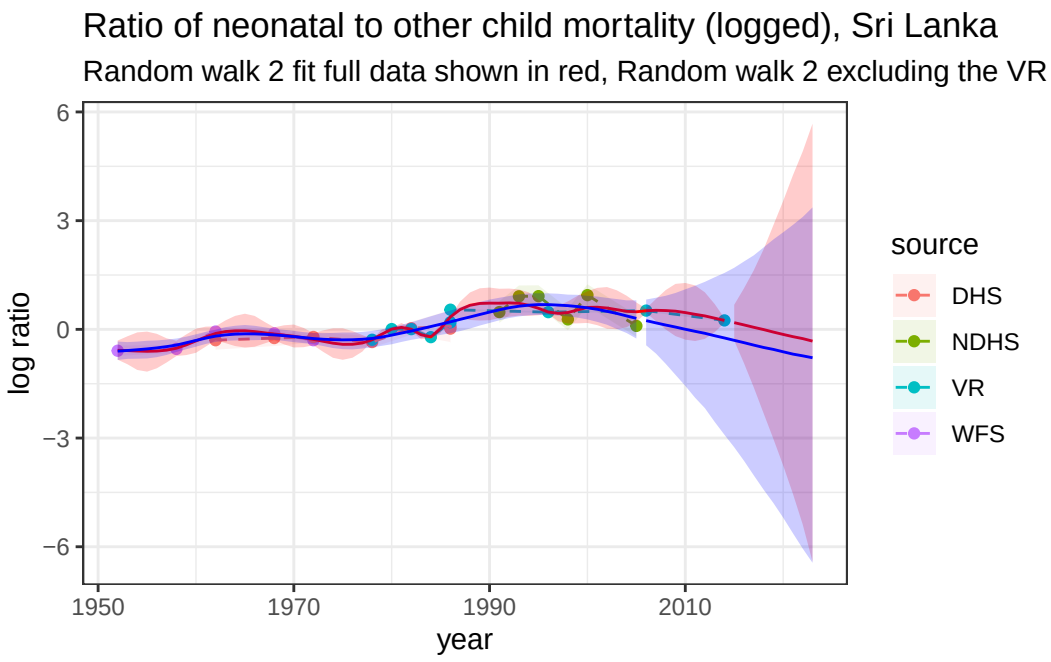
```
mod4 <- stan(data = stan_data2,
             file = here("code/models/lab10_3.stan"))
```

```
res4 <- mod4 %>%
  gather_draws(mu[t]) %>%
  median_qi() %>%
  mutate(year = years[t])

res4_p <- mod4 %>%
  gather_draws(mu_p[p]) %>%
  median_qi() %>%
  mutate(year = years[nyears]+p)

ggplot(lka2, 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 = res4, aes(year, .value)) +
  geom_ribbon(data = res4, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2)+
geom_line(data = res4_p, aes(year, .value),col="red") +
  geom_ribbon(data = res4_p, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2,fi
       y = "log ratio", subtitle = "Random walk2  fit excluding the VR data shown in black
```



Comparing

```
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 = res3, aes(year, .value),col="red") +
  geom_ribbon(data = res3, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2,fill
geom_line(data = res3_p, aes(year, .value),col="red") +
  geom_ribbon(data = res3_p, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2,fi
  geom_line(data = res4, aes(year, .value),col="blue") +
  geom_ribbon(data = res4, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2,fill
geom_line(data = res4_p, aes(year, .value),col="blue") +
  geom_ribbon(data = res4_p, aes(y = .value, ymin = .lower, ymax = .upper), alpha = 0.2,fi
       y = "log ratio", subtitle = "Random walk 2 fit full data shown in red, Random walk
```



Without VR data, we have no data after 2005. From the plot, we can see that without VR data, the confidence interval of the RW2 model is smoother than that of the model with VR data. Also, the projections have similar trend, the difference is that the projection of RW2

model without VR data is smaller than the projection of RW2 model with VR data.

Question 6

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

Answer

I prefer the RW2 model with VR data. Comparing to linear and RW1 models, it has better fit and valid projections. And VR data should not be excluded without valid reasons, as it provides valuable information for 2005-2014 and influences the projection.