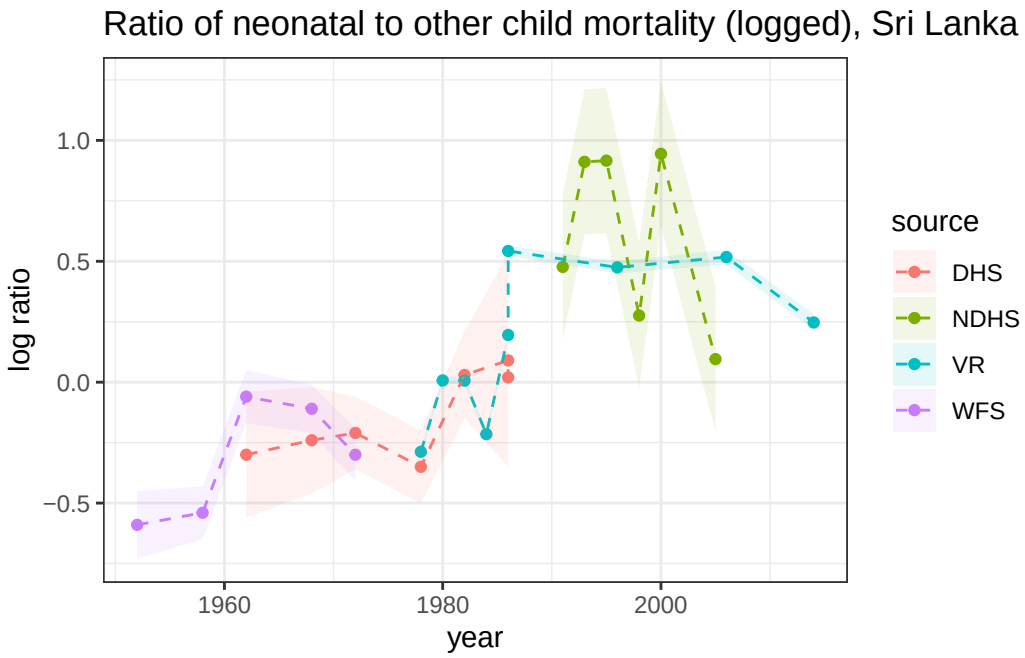
Week 10: Temporal data

25/03/23

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("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")
```



Fitting a linear model

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

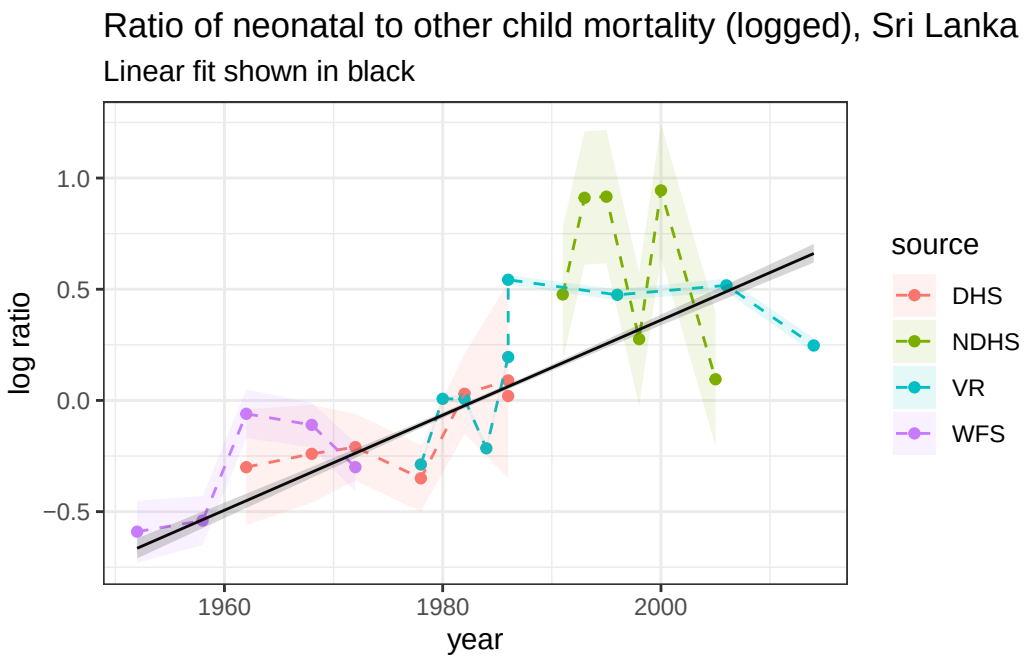
```
observed_years <- lka$year
years <- min(observed_years):max(observed_years)
nyears <- length(years)

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("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")
```



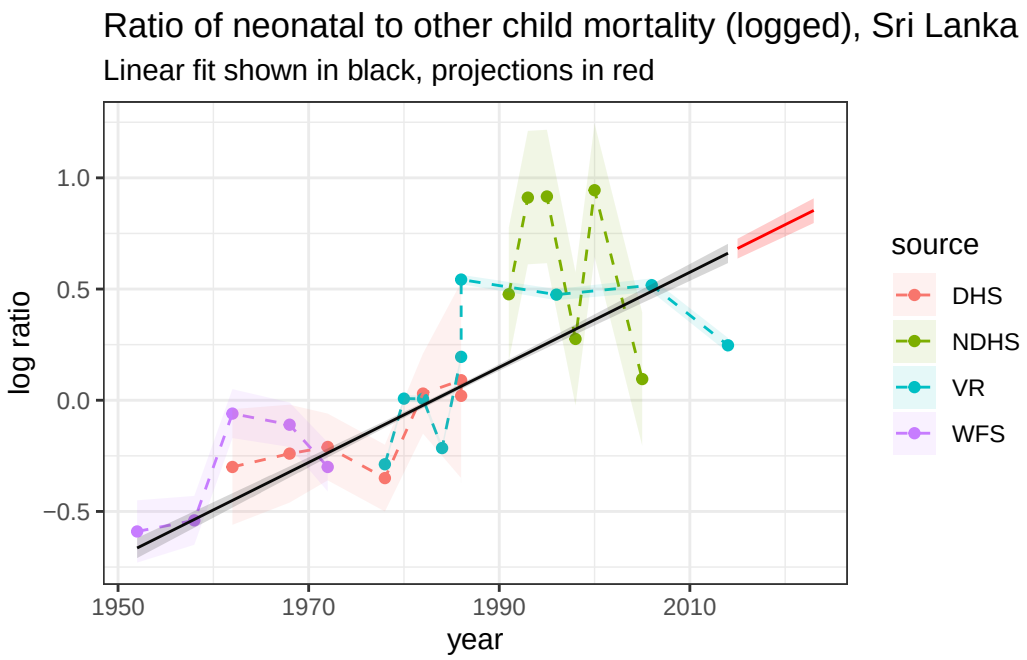
Question 1

```
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("lka2.stan"))
mod2
```

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

res_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, 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 = "Linear fit shown in black, projections in red")
```



Random walks

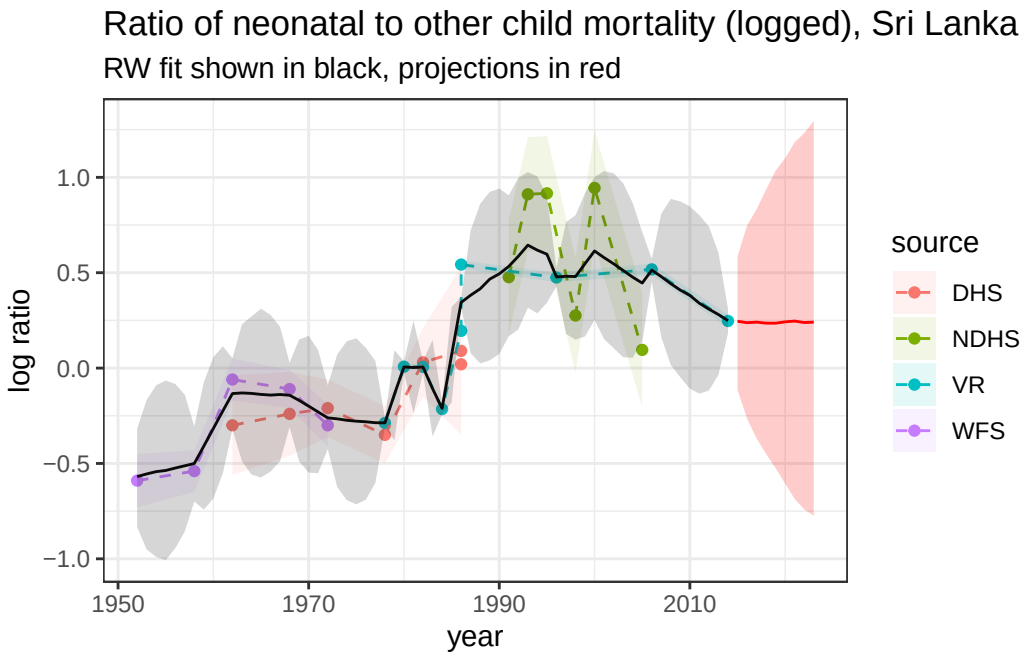
Question 2

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

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

rwres_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 = rwres, aes(year, .value)) +
  geom_ribbon(data = rwres, aes(y = .value, ymin = .lower, ymax = .upper),
              alpha = 0.2)+
  geom_line(data = rwres_p, aes(year, .value),col="red") +
  geom_ribbon(data = rwres_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")
```

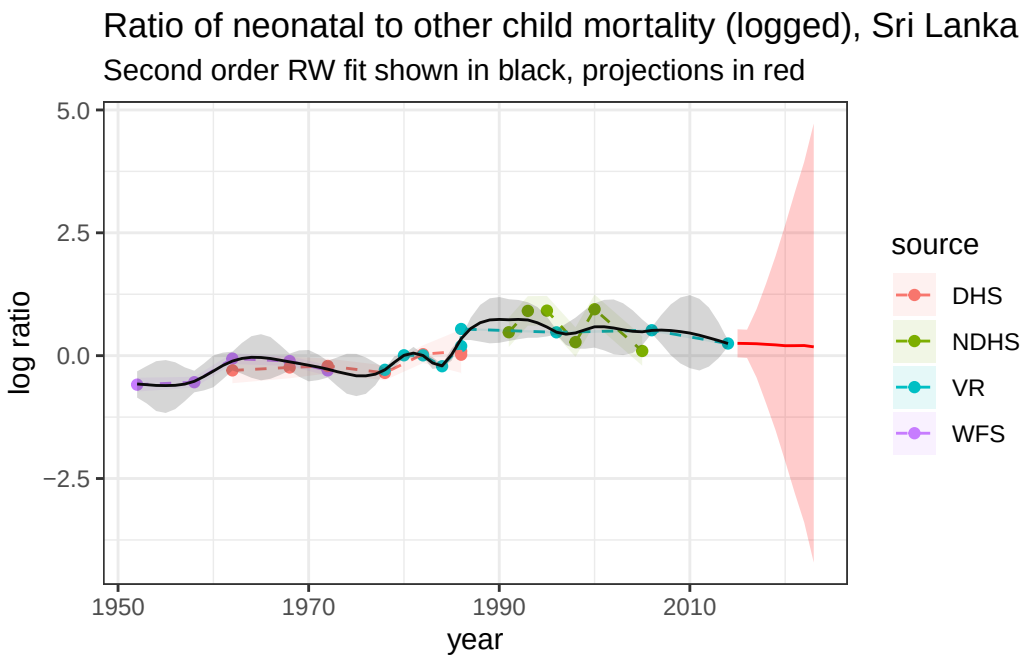


Question 3

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

```
rwres2=mod4 %>%
  gather_draws(mu[t]) %>%
  median_qi() %>%
  mutate(year=years[t])

rwres_p2=mod4 %>%
  gather_draws(mu_p[p]) %>%
  median_qi() %>%
  mutate(year=years[nyears]+p)
```

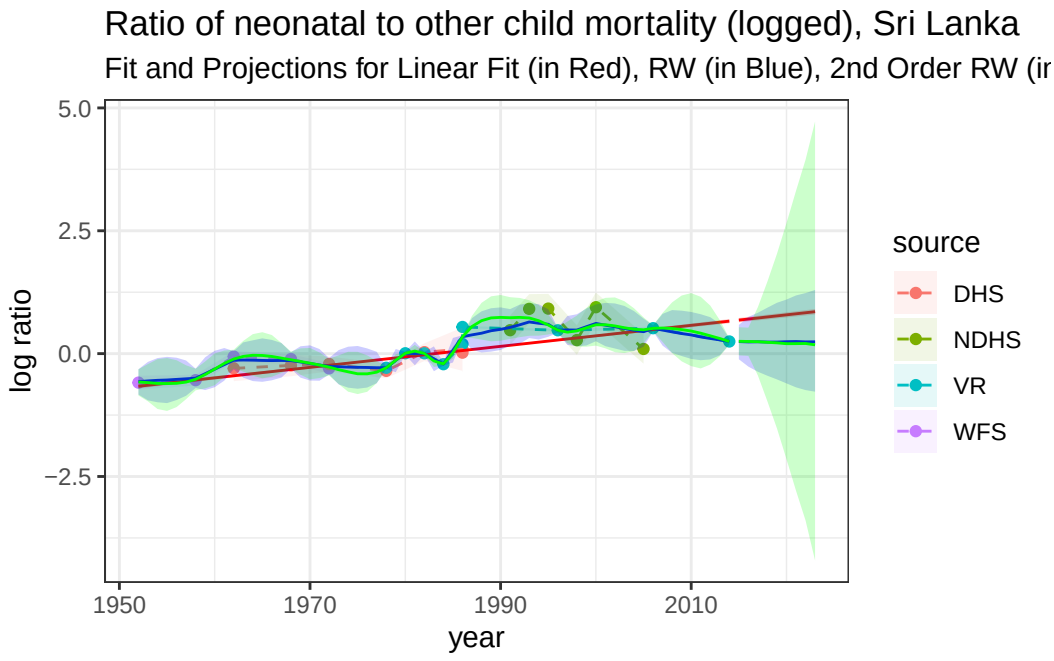
```
ggplot(lka, aes(year, logit_ratio)) +
  geom_point(aes( color = source)) +
  geom_line(aes( color = source), lty = 2) +
  geom_ribbon(aes(ymin = logit_ratio - se,
                  ymax = logit_ratio + se,
                  fill =  source), alpha = 0.1) +
  theme_bw()+
  geom_line(data = rwres2, aes(year, .value)) +
  geom_ribbon(data = rwres2, aes(y = .value, ymin = .lower, ymax = .upper),
              alpha = 0.2)+
  geom_line(data = rwres_p2, aes(year, .value),col="red") +
  geom_ribbon(data = rwres_p2, 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 = "Second order RW fit shown in black, projections in red")
```



Question 4

```
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),col="red") +
  geom_ribbon(data = res, aes(y = .value, ymin = .lower, ymax = .upper),
              fill="red", 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)+

  geom_line(data = rwres, aes(year, .value),col="blue") +
  geom_ribbon(data = rwres, aes(y = .value, ymin = .lower, ymax = .upper),
              fill="blue", alpha = 0.2)+
  geom_line(data = rwres_p, aes(year, .value),col="blue") +
  geom_ribbon(data = rwres_p, aes(y = .value, ymin = .lower, ymax = .upper),
              fill="blue", alpha = 0.2)+
```

```
geom_line(data = rwres2, aes(year, .value),col="green") +
geom_ribbon(data = rwres2, aes(y = .value, ymin = .lower, ymax = .upper),
            fill="green", alpha = 0.2)+
geom_line(data = rwres_p2, aes(year, .value),col="green") +
geom_ribbon(data = rwres_p2, aes(y = .value, ymin = .lower, ymax = .upper),
            fill="green", alpha = 0.2)+

theme_bw()+
labs(title = "Ratio of neonatal to other child mortality (logged), Sri Lanka",
     y = "log ratio",
     subtitle = "Fit and Projections for Linear Fit (in Red), RW (in Blue), 2nd Order RW
```



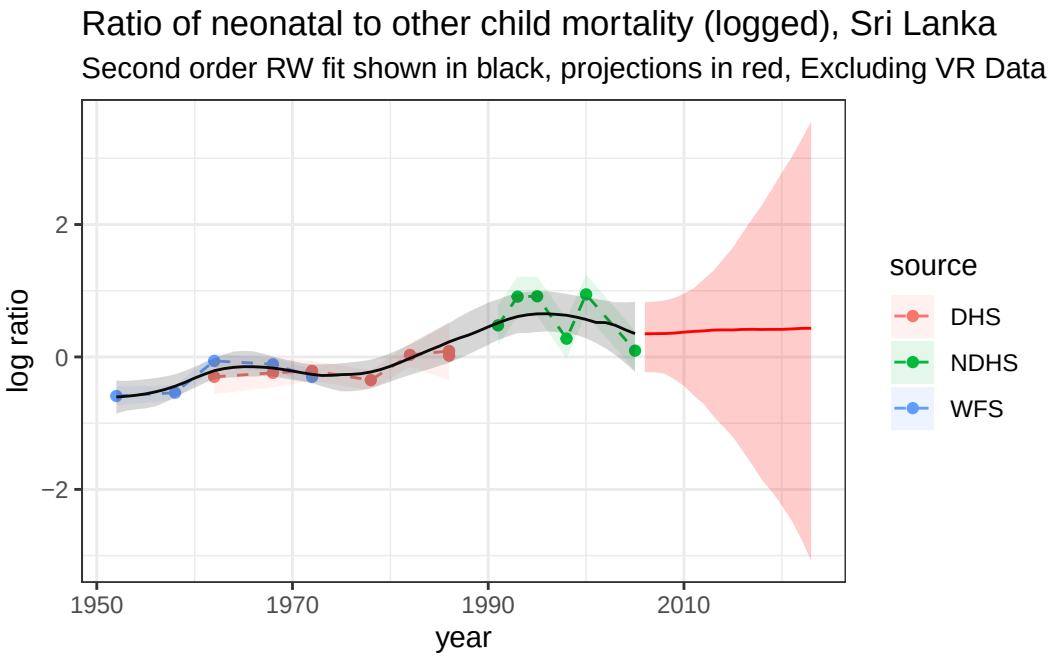
Question 5

```
lka2=lka %>% filter(source!="VR")
observed_years <- lka2$year
years <- min(observed_years):max(observed_years)
nyears <- length(years)
stan_data <- list(y = lka2$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_data,
             file = here("lka4.stan"),iter=2000)
```

```
rwres2=mod4 %>%
  gather_draws(mu[t]) %>%
  median_qi() %>%
  mutate(year=years[t])

rwres_p2=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 = rwres2, aes(year, .value)) +
  geom_ribbon(data = rwres2, aes(y = .value, ymin = .lower, ymax = .upper),
              alpha = 0.2)+
  geom_line(data = rwres_p2, aes(year, .value),col="red") +
  geom_ribbon(data = rwres_p2, 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 = "Second order RW fit shown in black, projections in red, Excluding VR Da
```



When the VR data is removed, there is less data available to fit the model. The latest year with data available is now 2005, which means we have to make predictions for 18 years to get to 2023 instead of of for 9 years starting from 2014. The VR data is also considered to be the "best" source of data, and we lose that when we remove it.

Question 6

The second order random walk model gives reasonable point estimates and better predictions of the change in ratio of neonatal to other child mortality over time than the other two models (since the linear model does not put more weight on more recent observations and the first-order random walk only considers the previous year).