Producing Figures

CKP

Data

```
library(tidyverse)
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.4.1     v purrr   1.0.1
v tibble  3.2.1     v dplyr   1.1.2
v tidyr   1.3.0     v stringr 1.5.0
v readr   2.1.3     v forcats 1.0.0
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

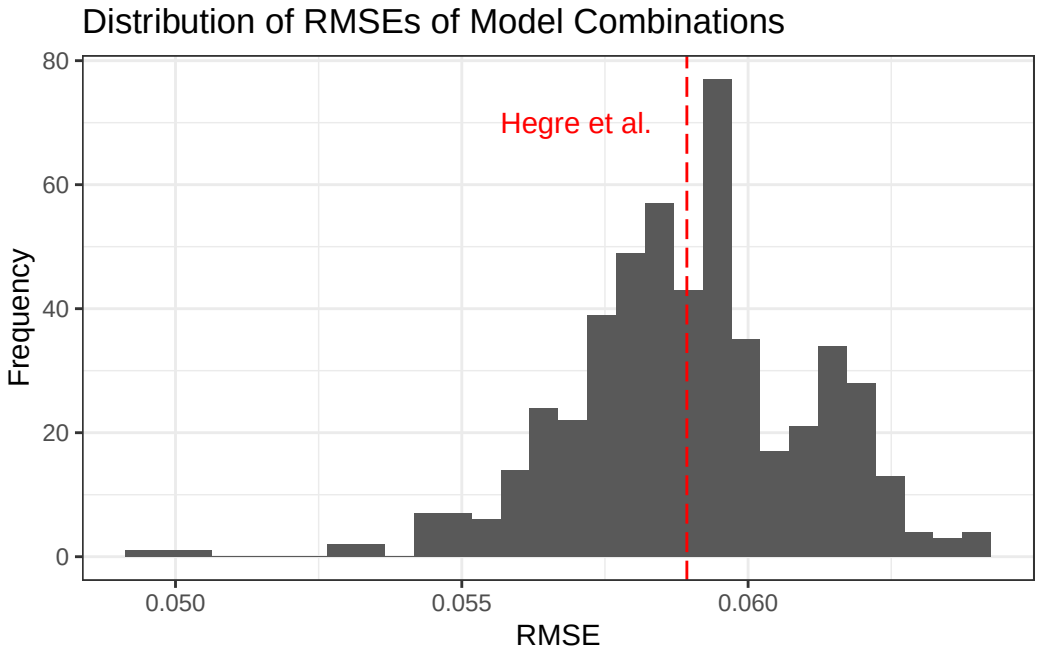
```
our_preds <- read.csv("our_predictions_for_all_years.csv")

real_conflict <- read.csv("real_prop_conflict.csv")
Hegre_preds <- read.csv("predictions_by_model.csv")
Hegre_preds <- Hegre_preds %>%
  group_by(year) %>%
  summarize(hegre_preds = mean(sh_w_c))


full_preds <- left_join(our_preds, real_conflict, by = join_by(year))

full_preds <- left_join(full_preds, Hegre_preds, by = join_by(year))

calc_errors <- full_preds %>%
  mutate(sq_error = (prop_conflict - predictions)^2,
         abs_error = abs(prop_conflict - predictions),
```

```
          Hegre = ifelse(combination == "(0, 1, 2, 3, 4, 5, 6, 7, 8)", 1, 0),
          lowest = ifelse(combination == "(0,)", 1, 0))

hegre_errors <- calc_errors %>%
  filter(Hegre == 1) %>%
  group_by(year) %>%
  mutate(hegre_error = abs_error) %>%
  select(year, hegre_error)

lowest_rmse <- calc_errors %>%
  filter(lowest == 1) %>%
  group_by(year) %>%
  mutate(lowest_error = abs_error) %>%
  select(year, lowest_error)


calc_errors <- left_join(calc_errors,hegre_errors, by = join_by(year))


calc_errors <- left_join(calc_errors,lowest_rmse, by = join_by(year))

rmses <- calc_errors %>%
  group_by(combination) %>%
  summarize(rmse = sqrt(mean(sq_error))) %>%
  mutate(Hegre = ifelse(combination == "(0, 1, 2, 3, 4, 5, 6, 7, 8)", 1, 0))
```

Creating Figure 1

Creating figure of the RMSEs for all of the models across the 2010 to 2018 period.

```
fig1 <- ggplot(rmses, aes(rmse)) +
  geom_histogram() +
  labs(title = "Distribution of RMSEs of Model Combinations", x = "RMSE", y = "Frequency")
  geom_vline(xintercept = 0.05893075, colour="red", linetype = "longdash") +
  annotate("text", x=.057, y=70, label= "Hegre et al.", col="red") +
  theme_bw()
fig1
```

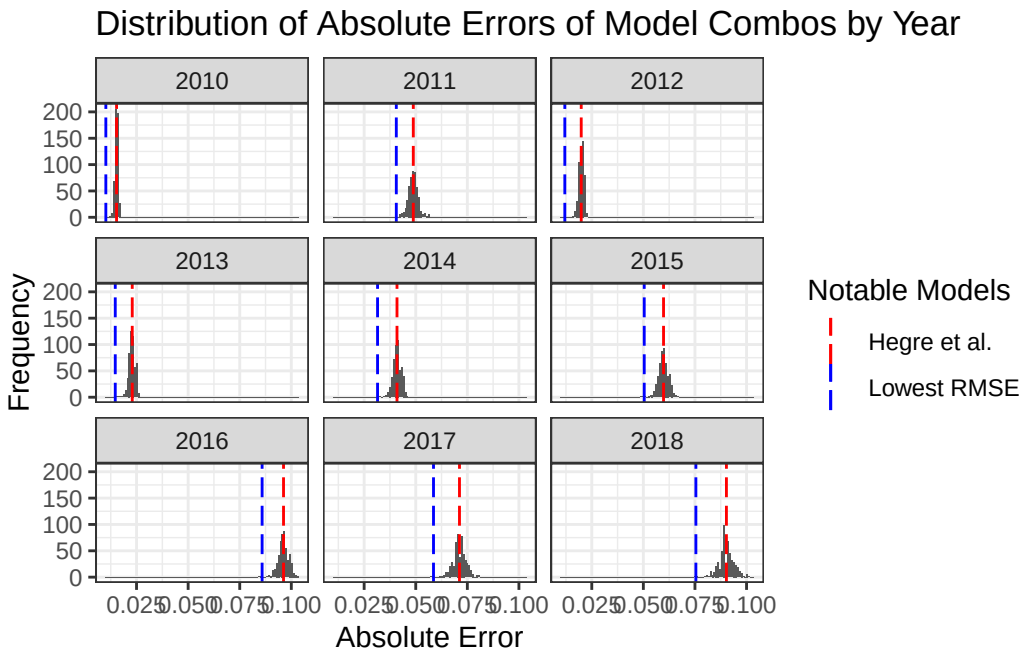



```
ggsave("fig1.png", width = 5, height = 5, dpi = 300)
```

Creating Figure 2

Creating the distribution of absolute error by year.

```
fig2 <- ggplot(calc_errors, aes(abs_error)) +
  geom_histogram(bins = 100) +
  labs(title = "Distribution of Absolute Errors of Model Combos by Year", x = "Absolute Er
  geom_vline(aes(xintercept=hegre_error, colour="Hegre et al."), linetype = "longdash") +
  geom_vline(aes(xintercept=lowest_error, colour="Lowest RMSE"), linetype = "longdash") +
  facet_wrap(~year) +
  scale_color_manual(name = "Notable Models", values = c("Lowest RMSE" = "blue", "Hegre et
  theme_bw()
fig2
```




```
ggsave("fig2.png", width = 10, height = 6, dpi = 300)
```

```
rmses <- rmses[order(rmses$rmse),]
```