# Brier Score

```r
library(dplyr)
library(ggplot2)
library(tidyr)
library(knitr)

load("data/HDA.RData")
load("~/GitHub/soccer-live-predictions/soccer-live-predictions/scrape/data/reds.RData")

nrow(HDA)
```
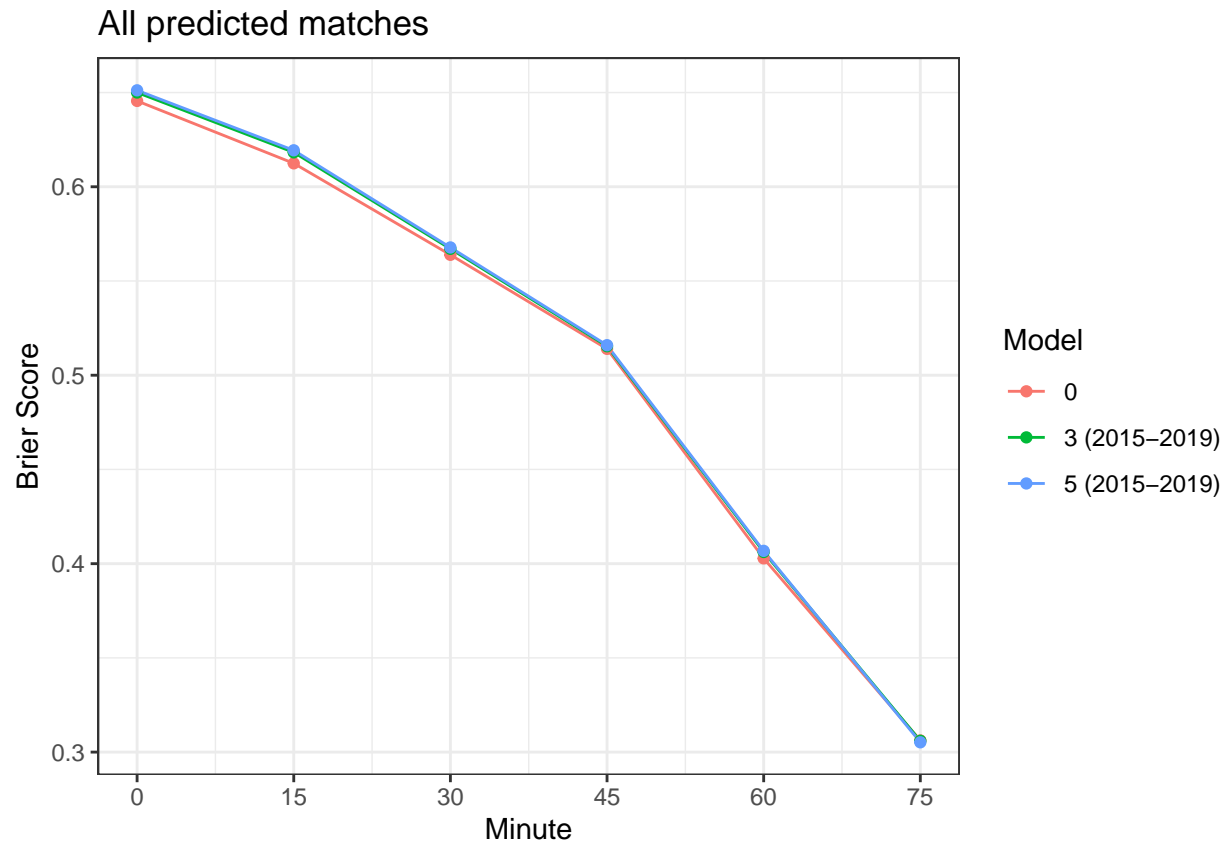
```
## [1] 333
```

```r
all = tibble(Brier = apply(HDA[,c(129:134, 147:158)], 2, mean),
             Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 3)),
             Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6),
                              rep("5 (2015-2019)", 6)),
                            levels = c("0", "3 (2015-2019)", "5 (2015-2019)")))
all %>%
  ggplot(aes(x = Minute, y = Brier, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches") +
  ylab("Brier Score")
```

## All predicted matches



```
all %>%
  pivot_wider(id_cols = "Model", values_from = "Brier", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```
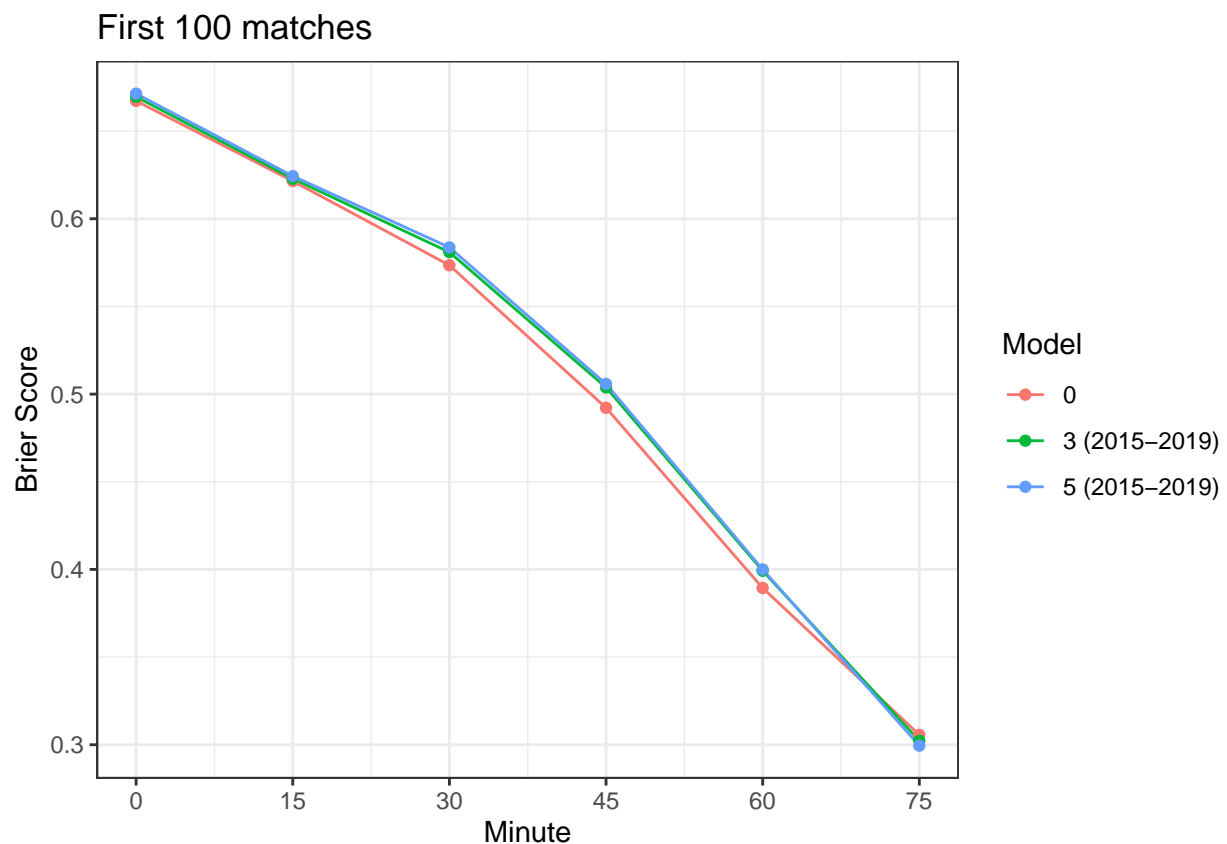
| Model | Minute 0 | Minute 15 | Minute 30 | Minute 45 | Minute 60 | Minute 75 |
|---|---|---|---|---|---|---|
| 0 | 0.6455559 | 0.6124371 | 0.5638907 | 0.5139790 | 0.4028096 | 0.3061762 |
| 3 (2015-2019) | 0.6500206 | 0.6182509 | 0.5670696 | 0.5153692 | 0.4061977 | 0.3059968 |
| 5 (2015-2019) | 0.6511736 | 0.6193479 | 0.5678061 | 0.5159370 | 0.4067885 | 0.3052059 |

```
first_100 = tibble(Brier = apply(HDA[c(1:100), c(129:134, 147:158)], 2, mean),
                    Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 3)),
                    Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6),
                                     rep("5 (2015-2019)", 6)),
                                   levels = c("0", "3 (2015-2019)", "5 (2015-2019)")))
first_100 %>%
  ggplot(aes(x = Minute, y = Brier, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("First 100 matches") +
  ylab("Brier Score")
```



First 100 matches

```
first_100 %>%
  pivot_wider(id_cols = "Model", values_from = "Brier", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```
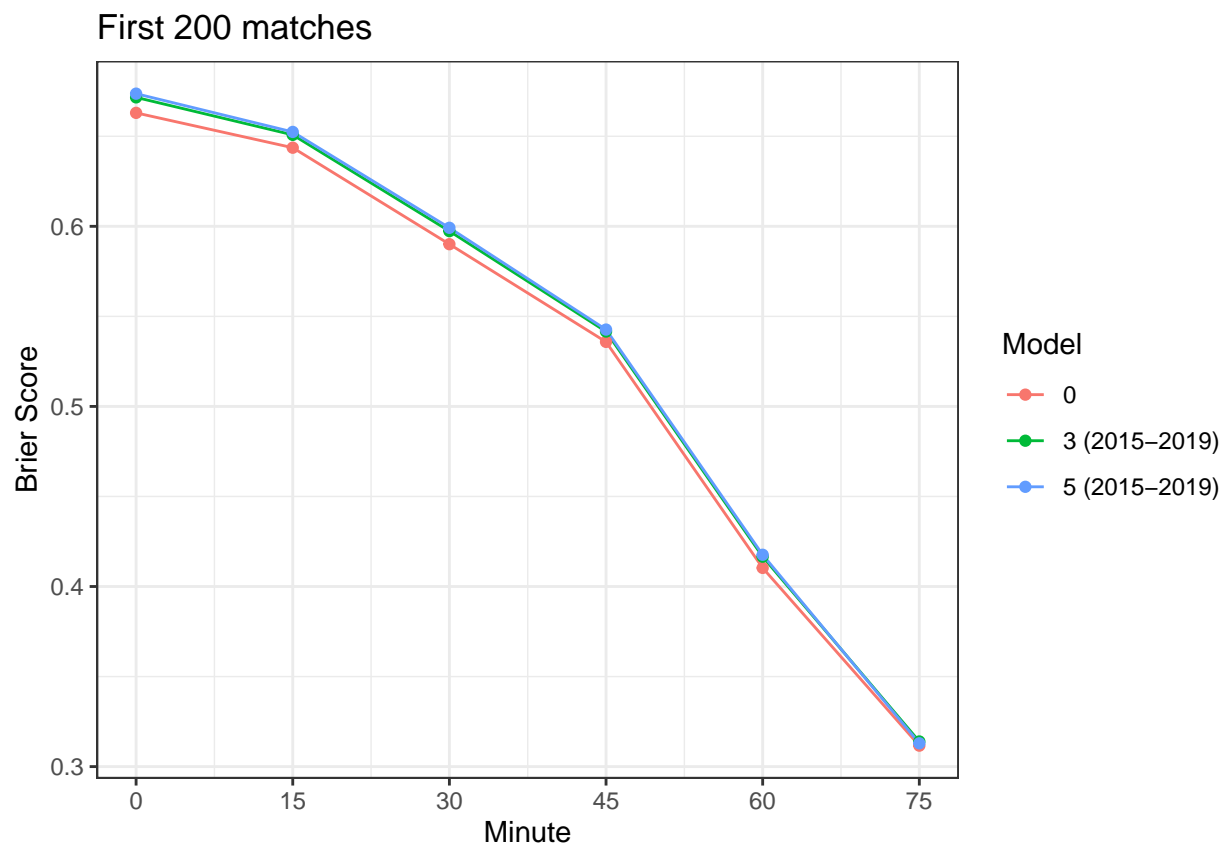
| Model | Minute 0 | Minute 15 | Minute 30 | Minute 45 | Minute 60 | Minute 75 |
|---|---|---|---|---|---|---|
| 0 | 0.6672148 | 0.6215355 | 0.5734965 | 0.4921946 | 0.3894027 | 0.3056346 |
| 3 (2015-2019) | 0.6695983 | 0.6228625 | 0.5809651 | 0.5038043 | 0.3992163 | 0.3022528 |
| 5 (2015-2019) | 0.6713621 | 0.6243722 | 0.5836960 | 0.5057874 | 0.3999847 | 0.2994870 |

```r
first_200 = tibble(Brier = apply(HDA[c(1:200), c(129:134, 147:158)], 2, mean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 3)),
                   Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6),
                                    rep("5 (2015-2019)", 6)),
                                  levels = c("0", "3 (2015-2019)", "5 (2015-2019)")))
first_200 %>%
  ggplot(aes(x = Minute, y = Brier, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("First 200 matches") +
  ylab("Brier Score")
```



First 200 matches

```r
first_200 %>%
  pivot_wider(id_cols = "Model", values_from = "Brier", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```
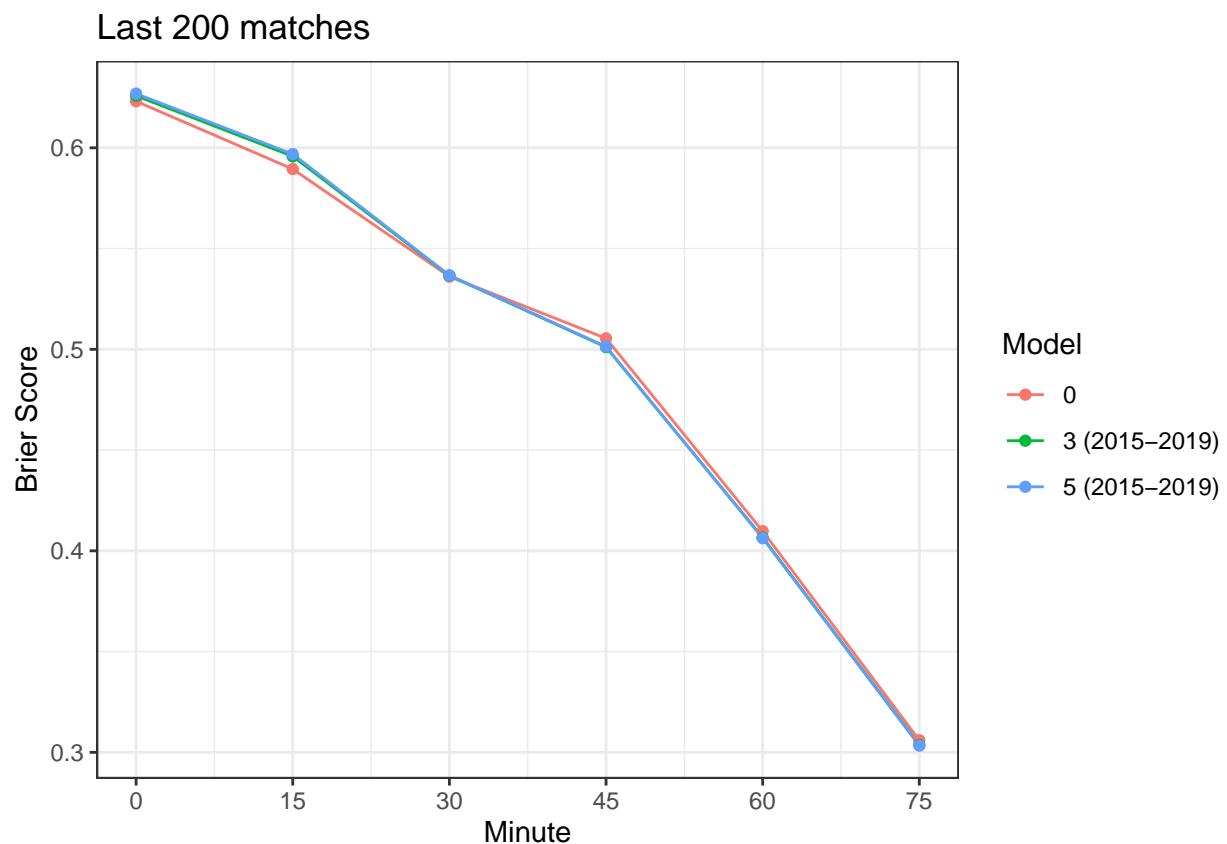
| Model | Minute 0 | Minute 15 | Minute 30 | Minute 45 | Minute 60 | Minute 75 |
|---|---|---|---|---|---|---|
| 0 | 0.6629896 | 0.6436404 | 0.5900750 | 0.5358331 | 0.4103278 | 0.3116472 |
| 3 (2015-2019) | 0.6715935 | 0.6507746 | 0.5973337 | 0.5415779 | 0.4165650 | 0.3139543 |
| 5 (2015-2019) | 0.6736495 | 0.6523998 | 0.5991689 | 0.5426644 | 0.4176221 | 0.3129503 |

```
last_200 = tibble(Brier = apply(HDA[c(134:333), c(129:134, 147:158)], 2, mean),
                  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 3)),
                  Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6),
                                   rep("5 (2015-2019)", 6)),
                                 levels = c("0", "3 (2015-2019)", "5 (2015-2019)")))
last_200 %>%
  ggplot(aes(x = Minute, y = Brier, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("Last 200 matches") +
  ylab("Brier Score")
```
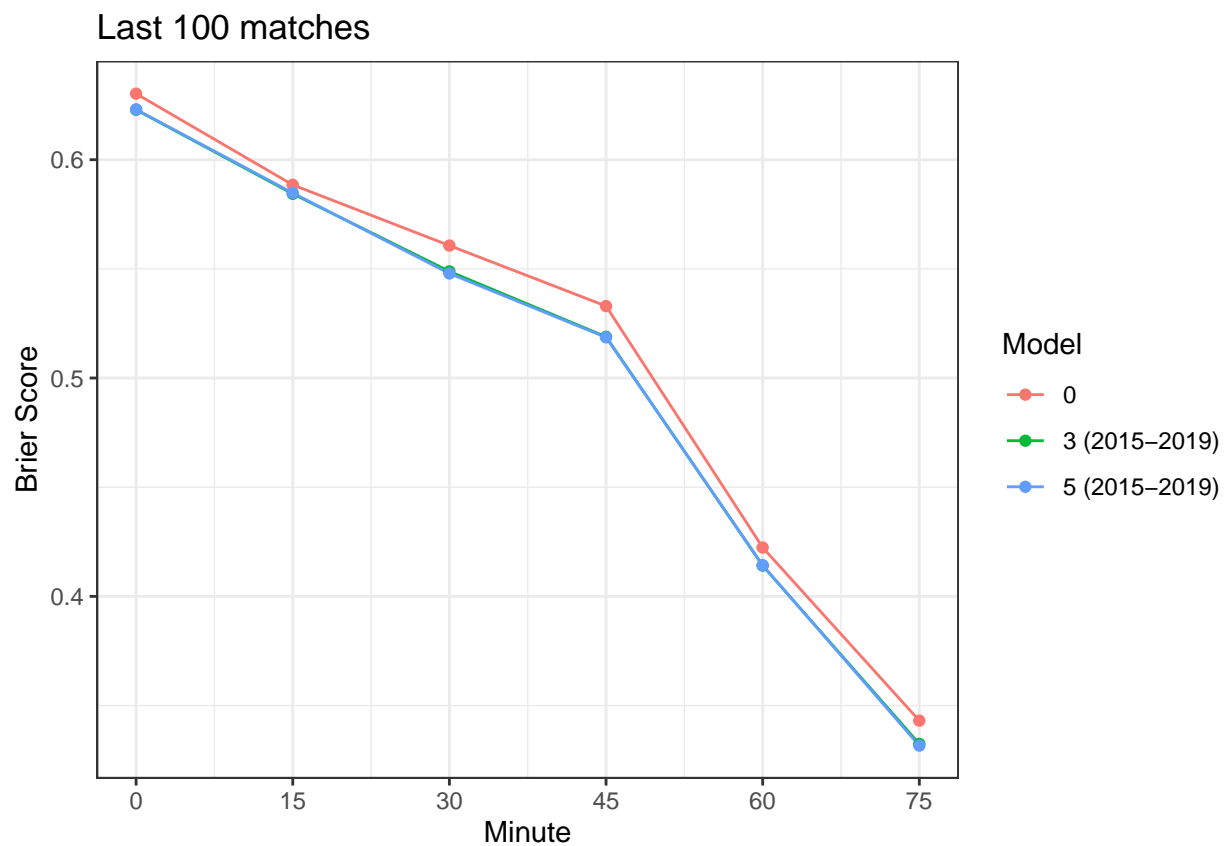


```
last_200 %>%
  pivot_wider(id_cols = "Model", values_from = "Brier", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

| Model | Minute 0 | Minute 15 | Minute 30 | Minute 45 | Minute 60 | Minute 75 |
|---|---|---|---|---|---|---|
| 0 | 0.6230999 | 0.5894508 | 0.5361184 | 0.5054219 | 0.4097503 | 0.3059723 |
| 3 (2015-2019) | 0.6258061 | 0.5957750 | 0.5365672 | 0.5010820 | 0.4066271 | 0.3036461 |
| 5 (2015-2019) | 0.6268681 | 0.5968787 | 0.5367094 | 0.5013498 | 0.4062767 | 0.3033166 |

```
last_100 = tibble(Brier = apply(HDA[c(234:333), c(129:134, 147:158)], 2, mean),
                  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 3)),
                  Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6),
                                   rep("5 (2015-2019)", 6)),
                                 levels = c("0", "3 (2015-2019)", "5 (2015-2019)"))))
last_100 %>%
  ggplot(aes(x = Minute, y = Brier, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("Last 100 matches") +
  ylab("Brier Score")
```



```
last_100 %>%
  pivot_wider(id_cols = "Model", values_from = "Brier", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

| Model | Minute 0 | Minute 15 | Minute 30 | Minute 45 | Minute 60 | Minute 75 |
|---|---|---|---|---|---|---|
| 0 | 0.6302431 | 0.5884844 | 0.5606849 | 0.5329531 | 0.4223720 | 0.3430799 |
| 3 (2015-2019) | 0.6228720 | 0.5843654 | 0.5488402 | 0.5188887 | 0.4140975 | 0.3324337 |
| 5 (2015-2019) | 0.6230258 | 0.5847853 | 0.5478892 | 0.5186324 | 0.4142637 | 0.3316354 |

```
matches = reds %>%
  filter(Season == 2020, Half == 1) %>%
  .$Match
length(matches)
```
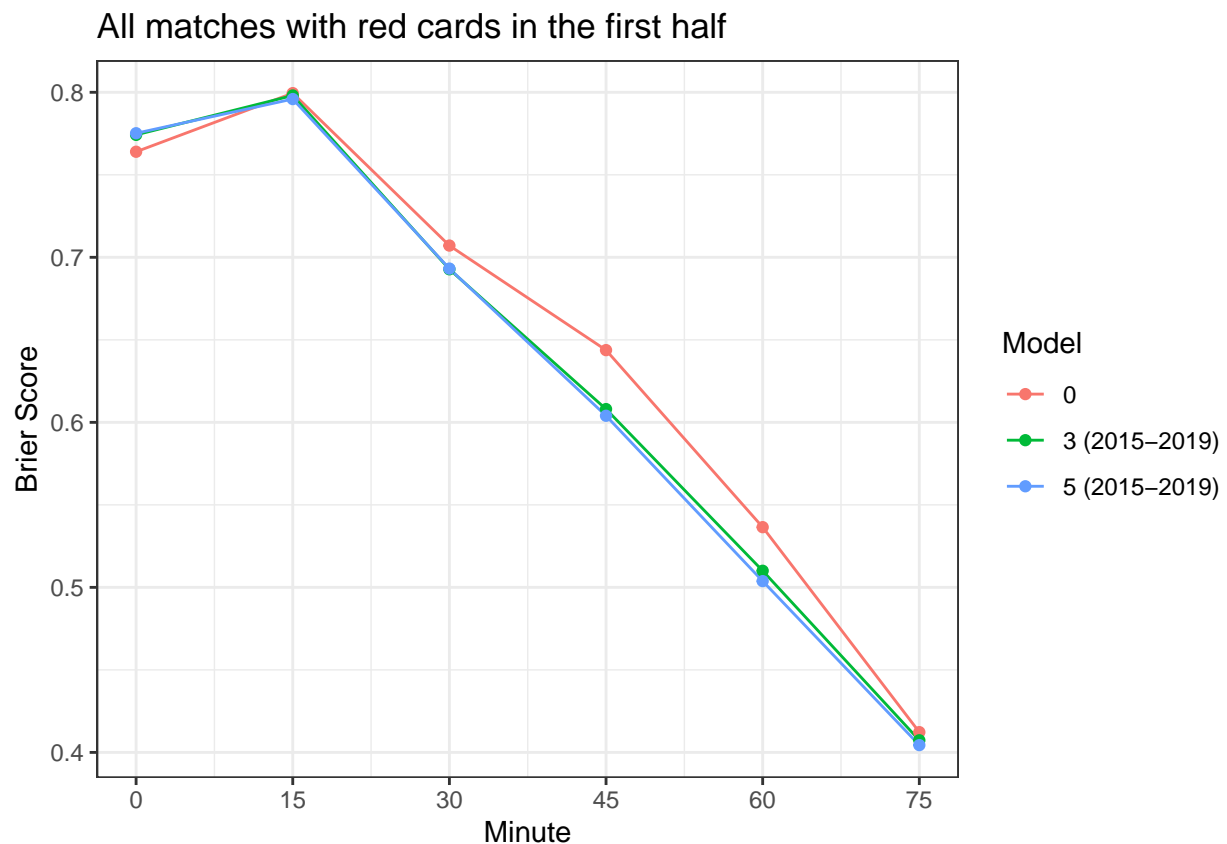
```
## [1] 23
```

```
HDA_reds = HDA %>%
  filter(Match %in% matches)

all_reds = tibble(Brier = apply(HDA_reds[,c(129:134, 147:158)], 2, mean),
                  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 3)),
                  Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6),
                                   rep("5 (2015-2019)", 6)),
                                 levels = c("0", "3 (2015-2019)", "5 (2015-2019)")))
all_reds %>%
  ggplot(aes(x = Minute, y = Brier, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All matches with red cards in the first half") +
  ylab("Brier Score")
```

```
all_reds %>%
  pivot_wider(id_cols = "Model", values_from = "Brier", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

| Model | Minute 0 | Minute 15 | Minute 30 | Minute 45 | Minute 60 | Minute 75 |
|-------|----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0.7639694 | 0.7994884 | 0.7071222 | 0.6437599 | 0.5364822 | 0.4122734 |
| 3 (2015-2019) | 0.7742053 | 0.7980646 | 0.6927592 | 0.6080206 | 0.5100461 | 0.4073383 |
| 5 (2015-2019) | 0.7751813 | 0.7959124 | 0.6931562 | 0.6039830 | 0.5037842 | 0.4044124 |