Brier Score

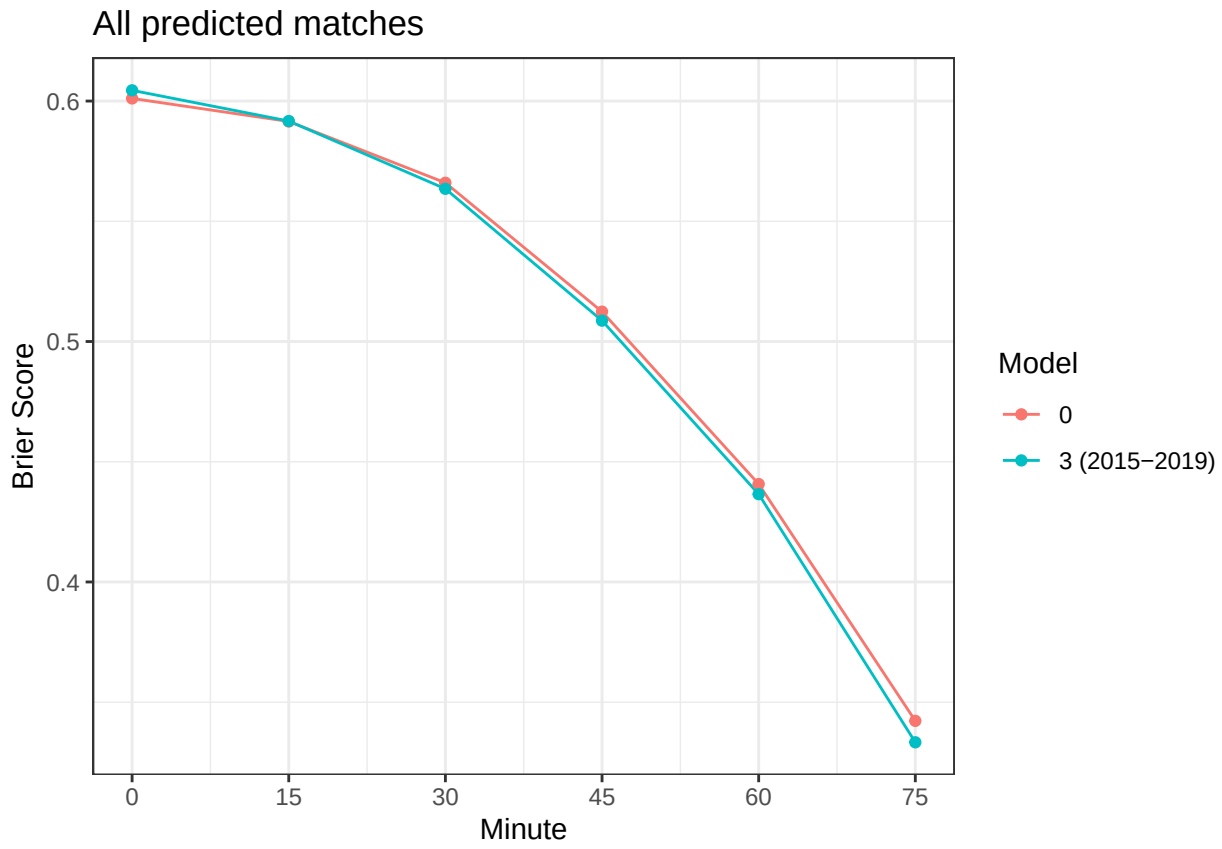
```
library(dplyr)
library(ggplot2)
library(tidyr)
library(knitr)

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

nrow(HDA)
```

```
## [1] 340
```

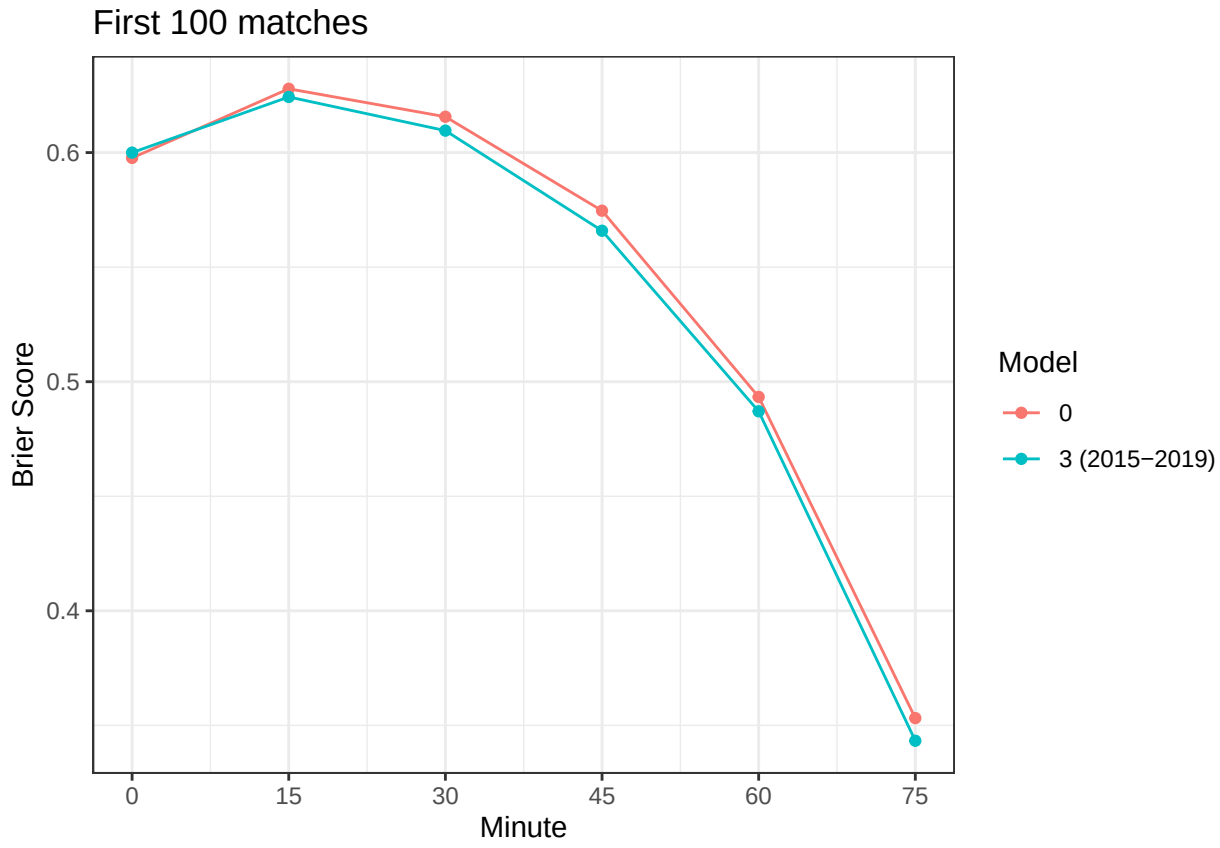
```
all = tibble(Brier = apply(HDA[,c(57:68)], 2, mean),
             Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
             Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                            levels = c("0", "3 (2015-2019)")))
all %>%
  ggplot(aes(x = Minute, y = Brier, 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 %>%
  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	0.6011055	0.5915227	0.5660078	0.5124139	0.4407243	0.3422376
3 (2015-2019)	0.6044661	0.5916919	0.5635232	0.5086914	0.4365300	0.3333226

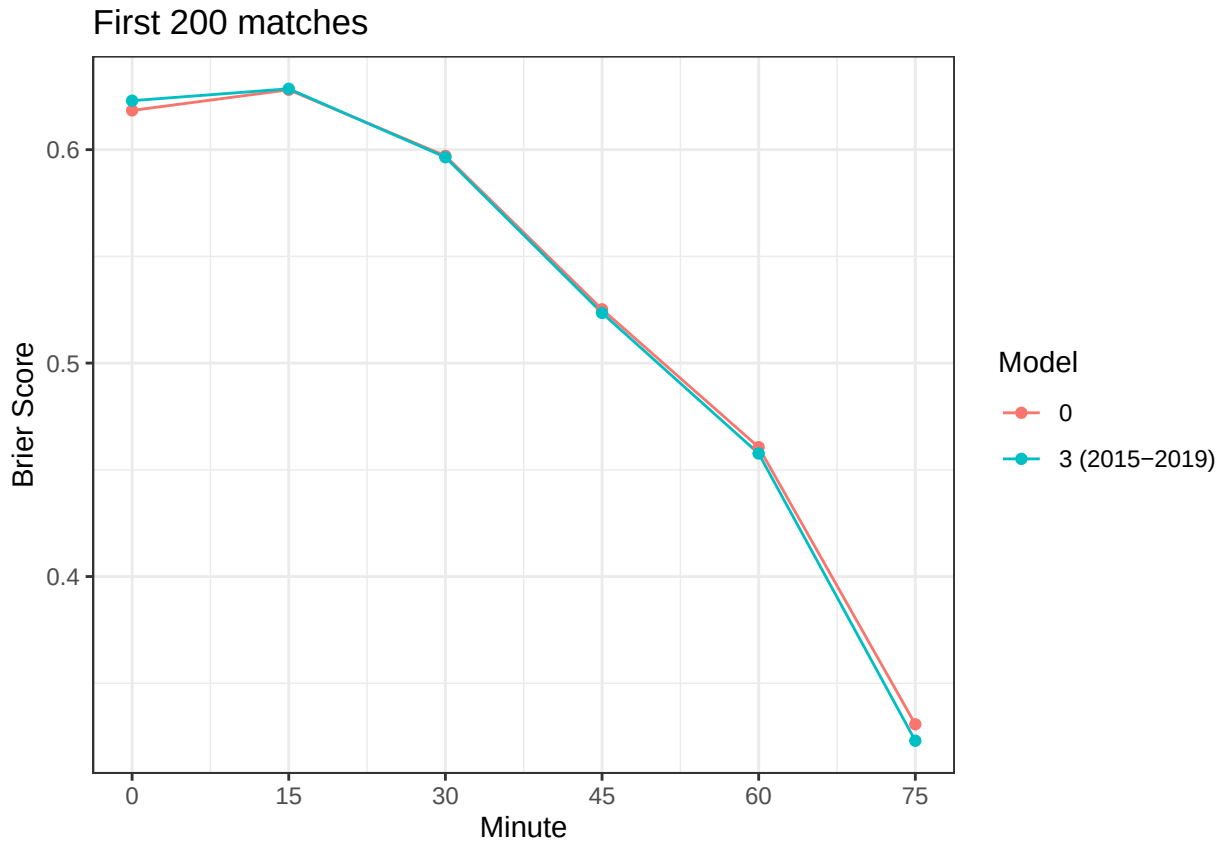
```
first_100 = tibble(Brier = apply(HDA[c(1:100), c(57:68)], 2, mean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                   Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                                  levels = c("0", "3 (2015-2019)")))
first_100 %>%
  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 %>%
  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	0.5976644	0.6278469	0.6156515	0.5746298	0.4933569	0.3531603
3 (2015-2019)	0.5999882	0.6242815	0.6095881	0.5658837	0.4870711	0.3432820

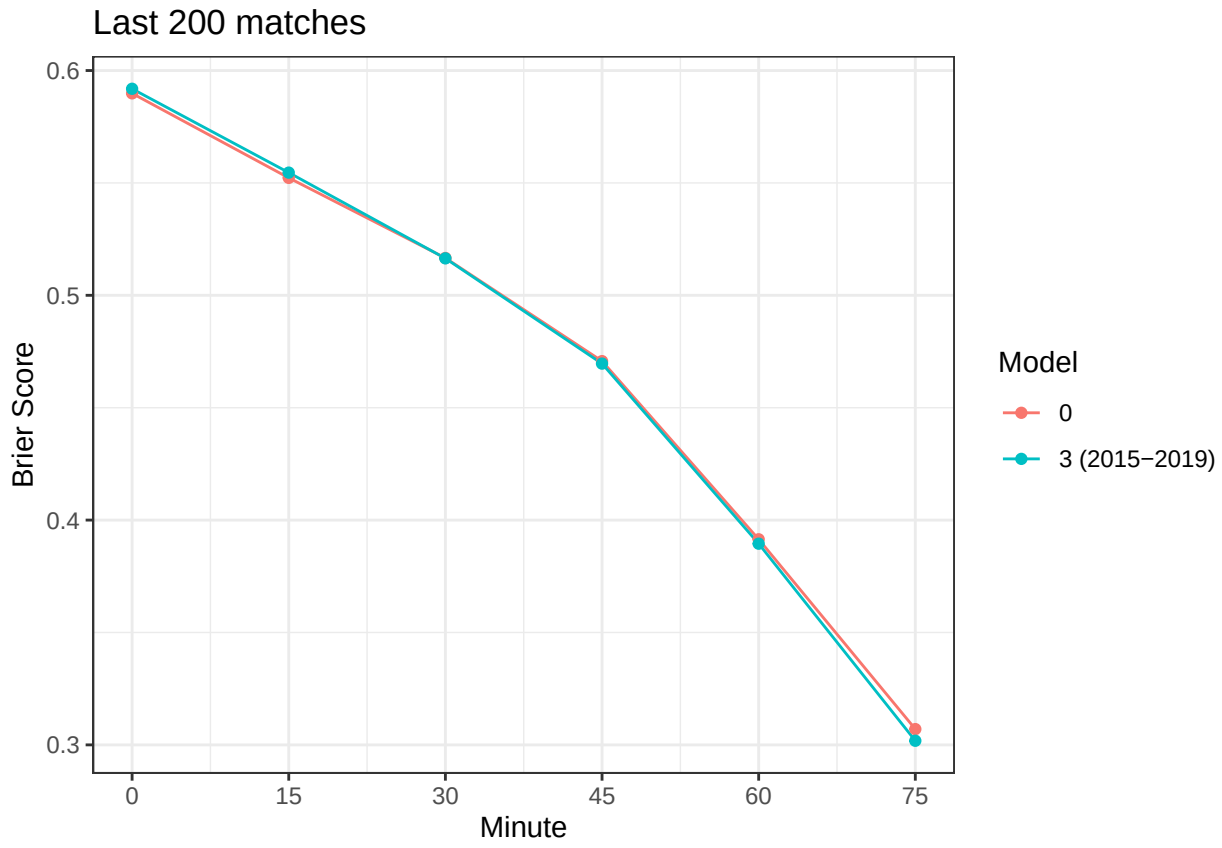
```
first_200 = tibble(Brier = apply(HDA[c(1:200), c(57:68)], 2, mean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                   Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                                  levels = c("0", "3 (2015-2019)")))
first_200 %>%
  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 %>%
  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	0.6183749	0.6280458	0.5971006	0.5252032	0.4606016	0.3307662
3 (2015-2019)	0.6229379	0.6285173	0.5964522	0.5234798	0.4576239	0.3230541

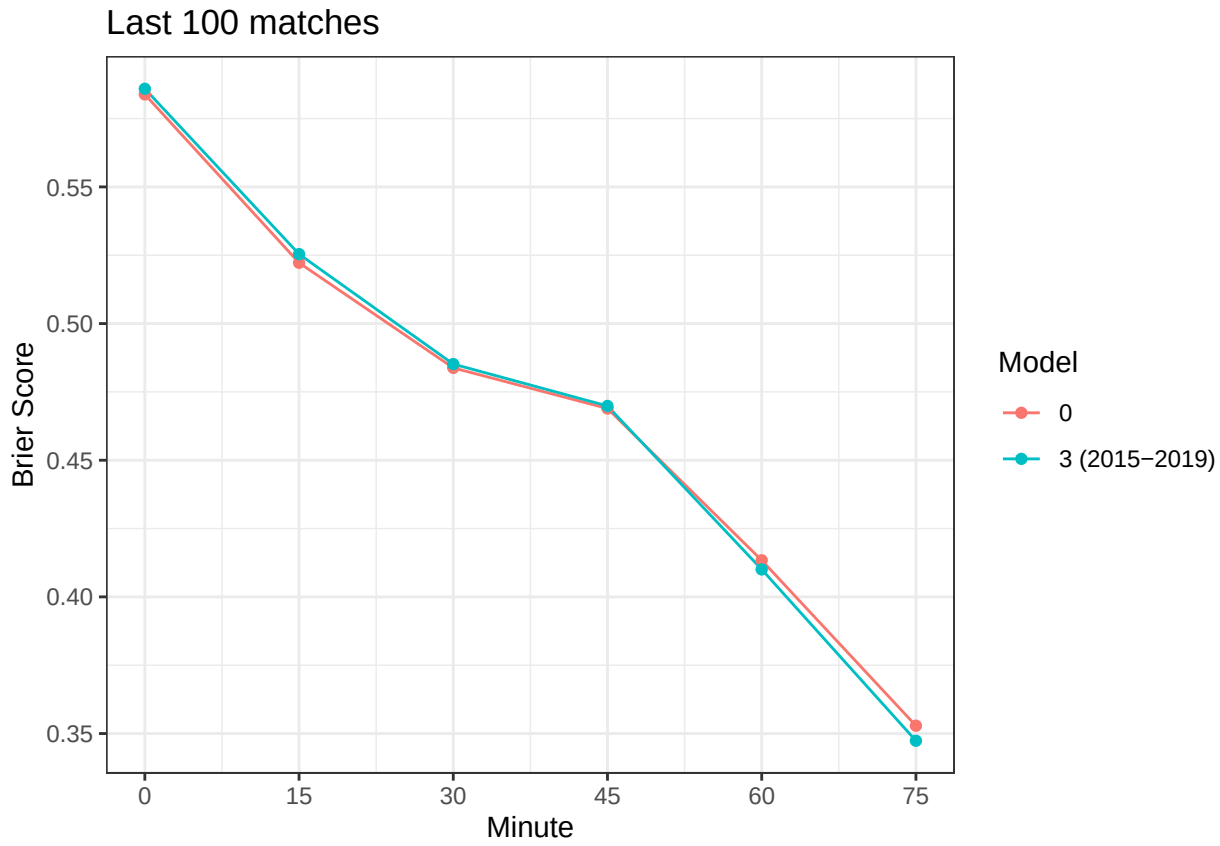
```
last_200 = tibble(Brier = apply(HDA[c(141:340), c(57:68)], 2, mean),
                  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                  Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                                 levels = c("0", "3 (2015-2019)")))
last_200 %>%
  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	0.5898486	0.5522471	0.5166005	0.4707222	0.3914555	0.3070150
3 (2015-2019)	0.5918764	0.5545814	0.5164374	0.4696099	0.3894978	0.3018292

```
last_100 = tibble(Brier = apply(HDA[c(241:340), c(57:68)], 2, mean),
                  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                  Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                                 levels = c("0", "3 (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()
```

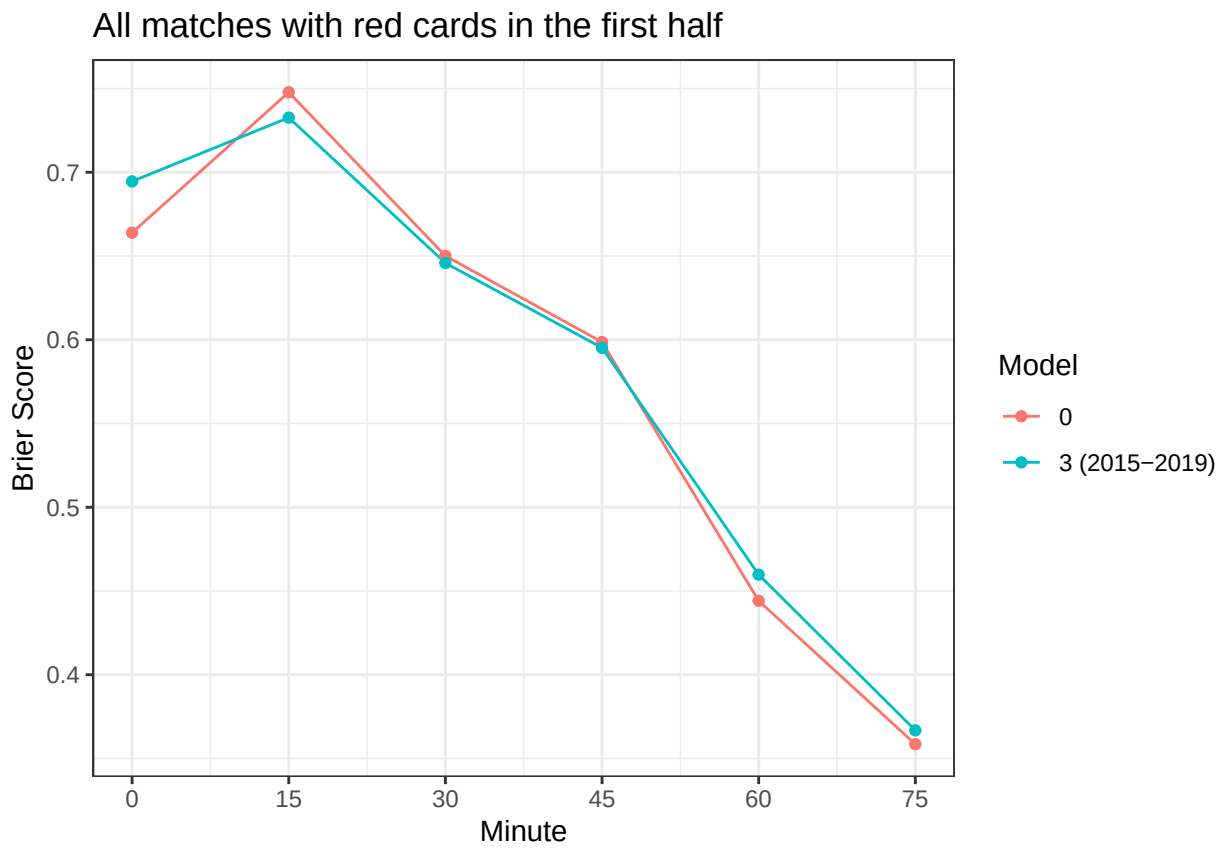
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.5838481	0.5222353	0.4837947	0.4689226	0.4133881	0.3528603
3 (2015-2019)	0.5858997	0.5254100	0.4851761	0.4698542	0.4100509	0.3473608

```
matches = reds %>%
  filter(Season == 2019, Half == 1) %>%
  .$Match
length(matches)
```

```
## [1] 17
```

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

all_reds = tibble(Brier = apply(HDA_reds[,c(57:68)], 2, mean),
                  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                  Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                                 levels = c("0", "3 (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	0.6639028	0.7477645	0.6500714	0.5986169	0.4441240	0.3585161
3 (2015-2019)	0.6945510	0.7326063	0.6458039	0.5952014	0.4597477	0.3668206