Brier Score

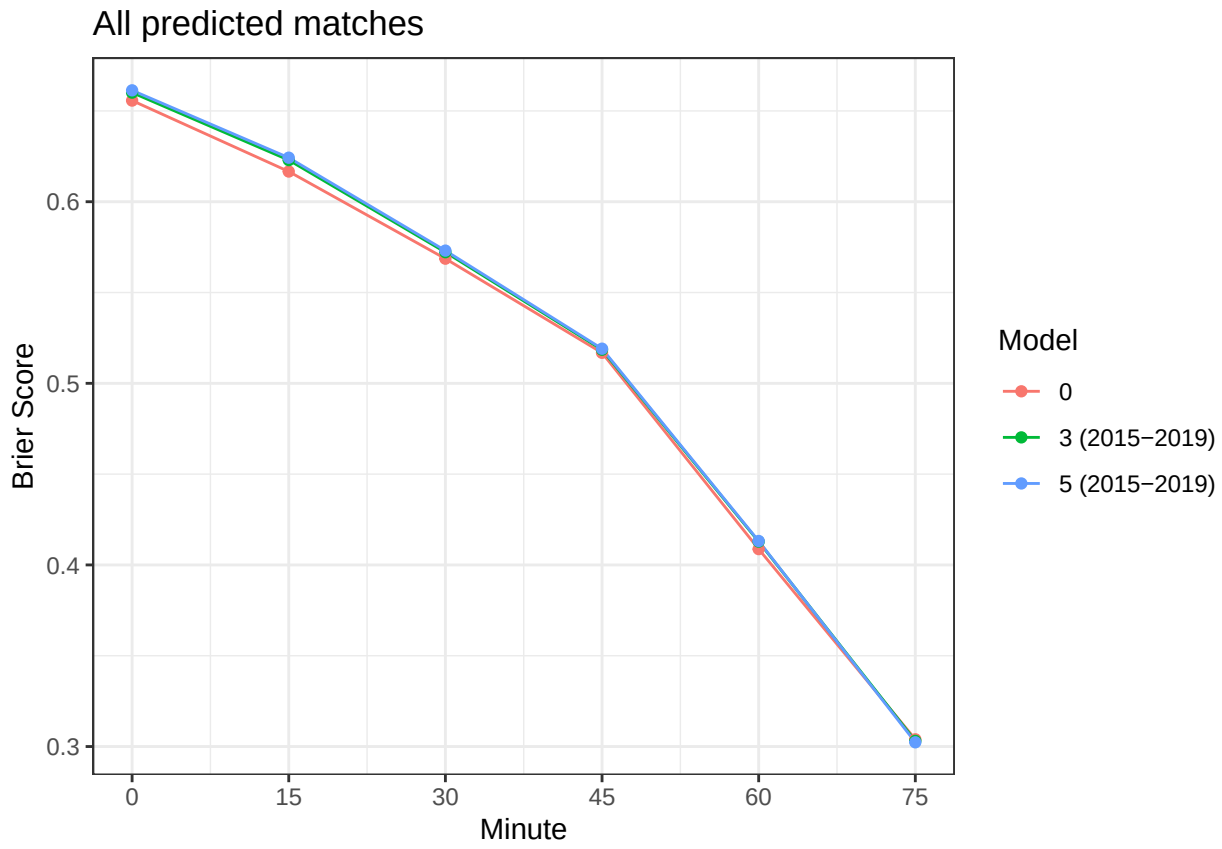
```
library(dplyr)
library(ggplot2)
library(tidyr)
library(knitr)

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

nrow(HDA)
```

```
## [1] 350
```

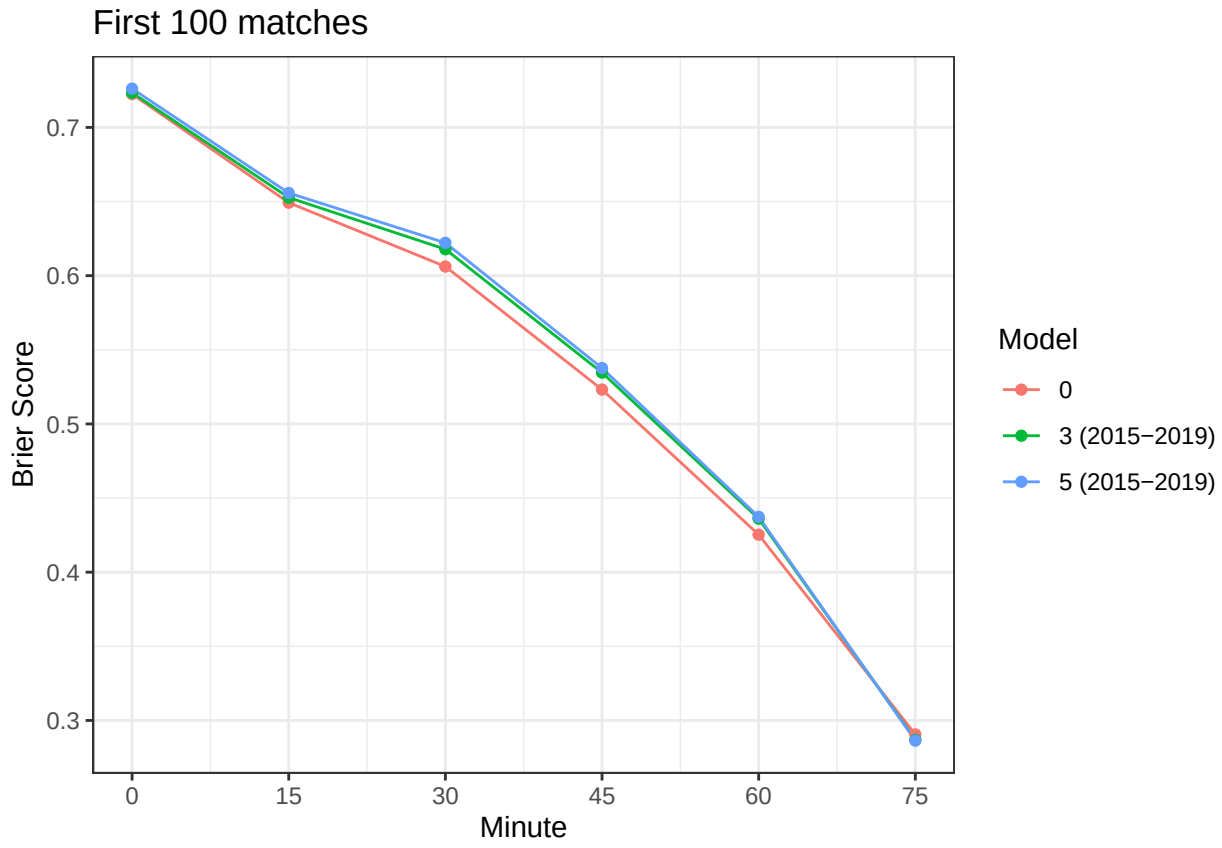
```
all = tibble(Brier = apply(HDA[,c(129:134, 147:158)], 2, mean),
             Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 3)),
             Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6),
                              rep("5 (2015-2019)", 6)),
                            levels = c("0", "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 %>%
  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	0.6557029	0.6166259	0.5686376	0.5168042	0.4087130	0.3039602
3 (2015-2019)	0.6600751	0.6229081	0.5721995	0.5185039	0.4128283	0.3031031
5 (2015-2019)	0.6613330	0.6242415	0.5730733	0.5190552	0.4131529	0.3023222

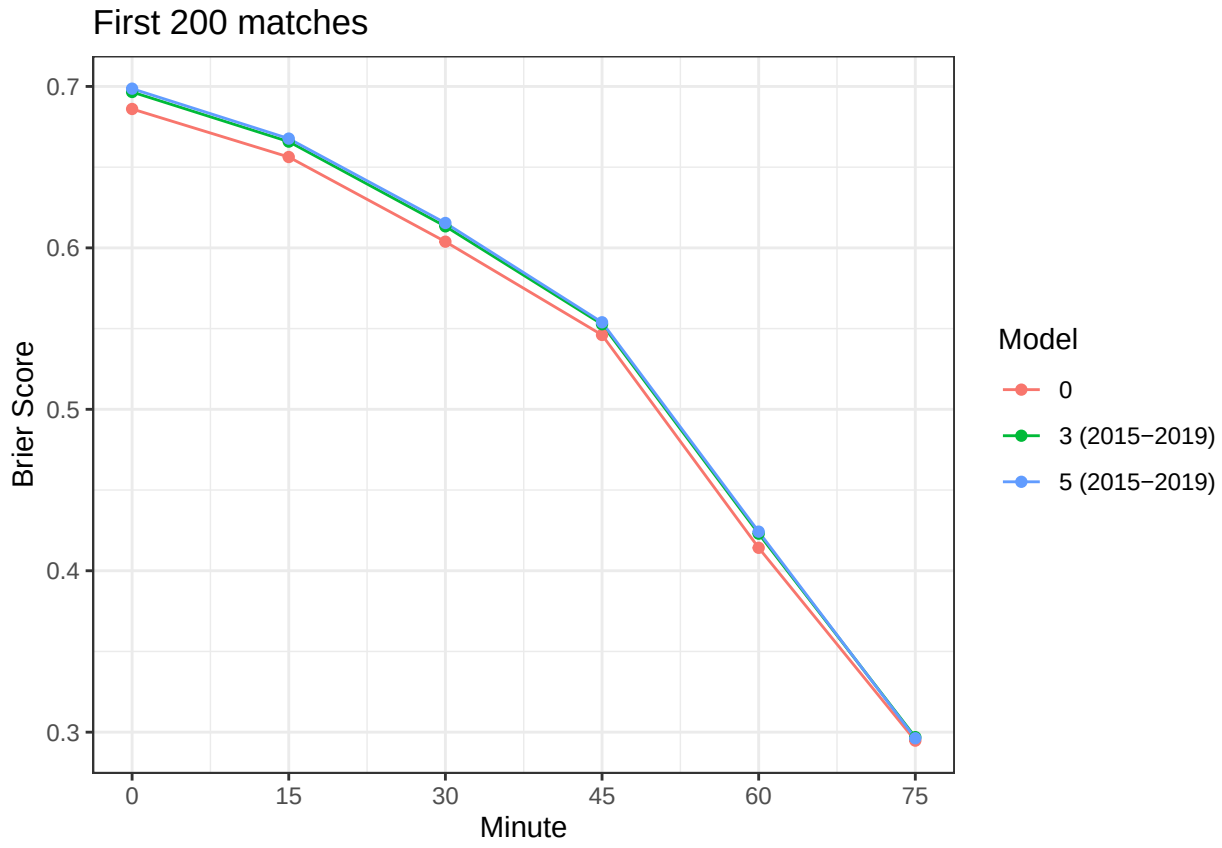
```
first_100 = tibble(Brier = apply(HDA[c(1:100), c(129:134, 147:158)], 2, mean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 3)),
                   Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6),
                                    rep("5 (2015-2019)", 6)),
                                  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 %>%
  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	0.7224852	0.6492050	0.6061532	0.5231855	0.4252740	0.2905962
3 (2015-2019)	0.7233623	0.6526449	0.6178094	0.5345832	0.4361013	0.2869517
5 (2015-2019)	0.7261112	0.6557405	0.6221257	0.5376784	0.4373227	0.2863898

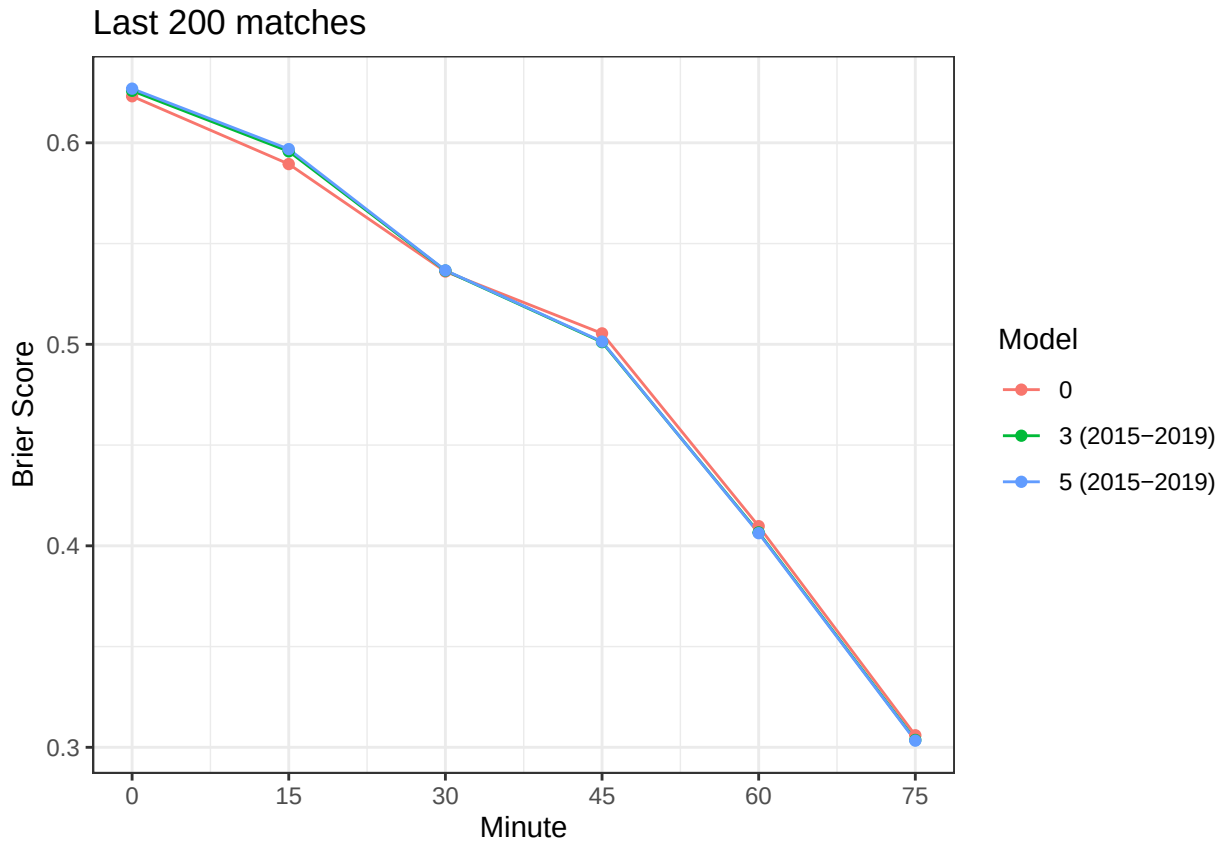
```
first_200 = tibble(Brier = apply(HDA[c(1:200), c(129:134, 147:158)], 2, mean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 3)),
                   Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6),
                                    rep("5 (2015-2019)", 6)),
                                  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 %>%
  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	0.6860374	0.6562795	0.6038389	0.5460537	0.4142008	0.2947619
3 (2015-2019)	0.6965541	0.6658075	0.6133372	0.5526954	0.4229632	0.2969022
5 (2015-2019)	0.6986174	0.6676894	0.6154140	0.5538981	0.4242134	0.2962528

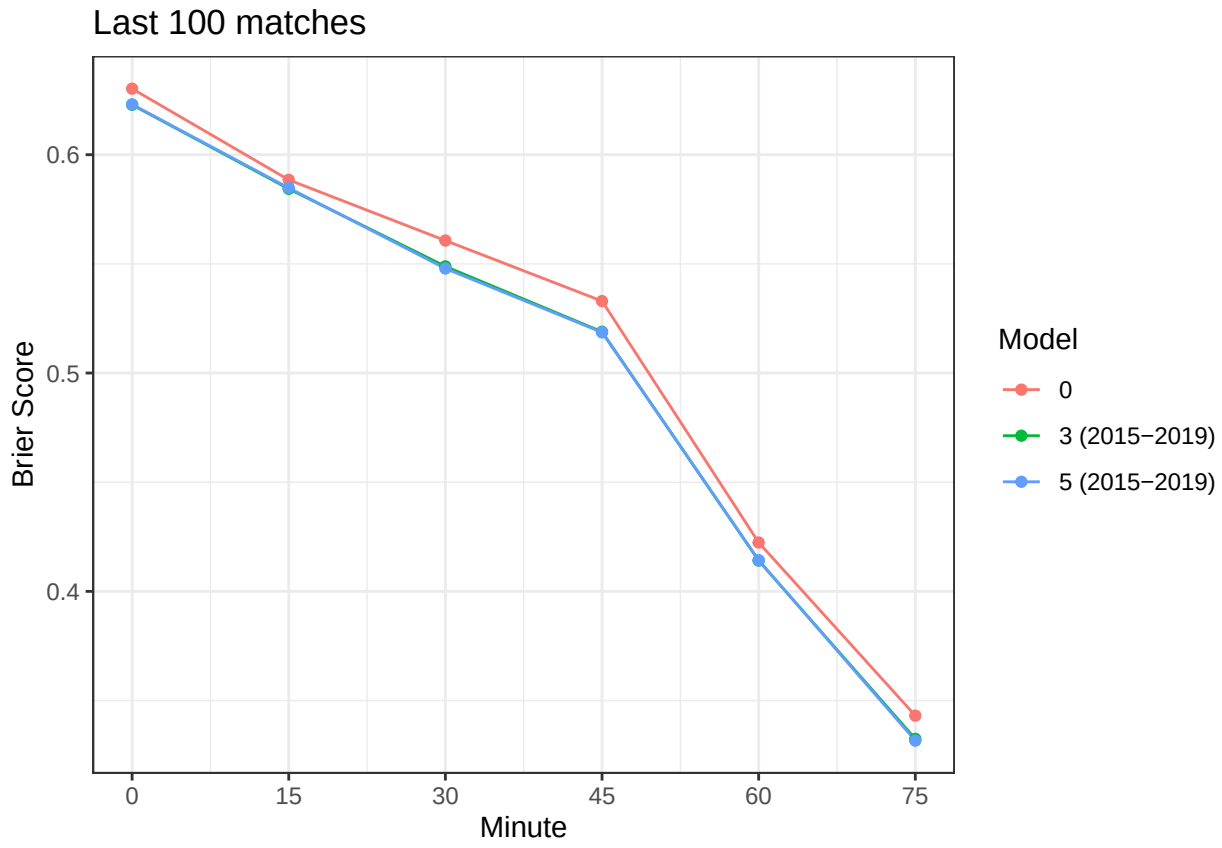
```
last_200 = tibble(Brier = apply(HDA[c(151:350), c(129:134, 147:158)], 2, mean),
                  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 3)),
                  Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6),
                                   rep("5 (2015-2019)", 6)),
                                 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(251:350), 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()
```

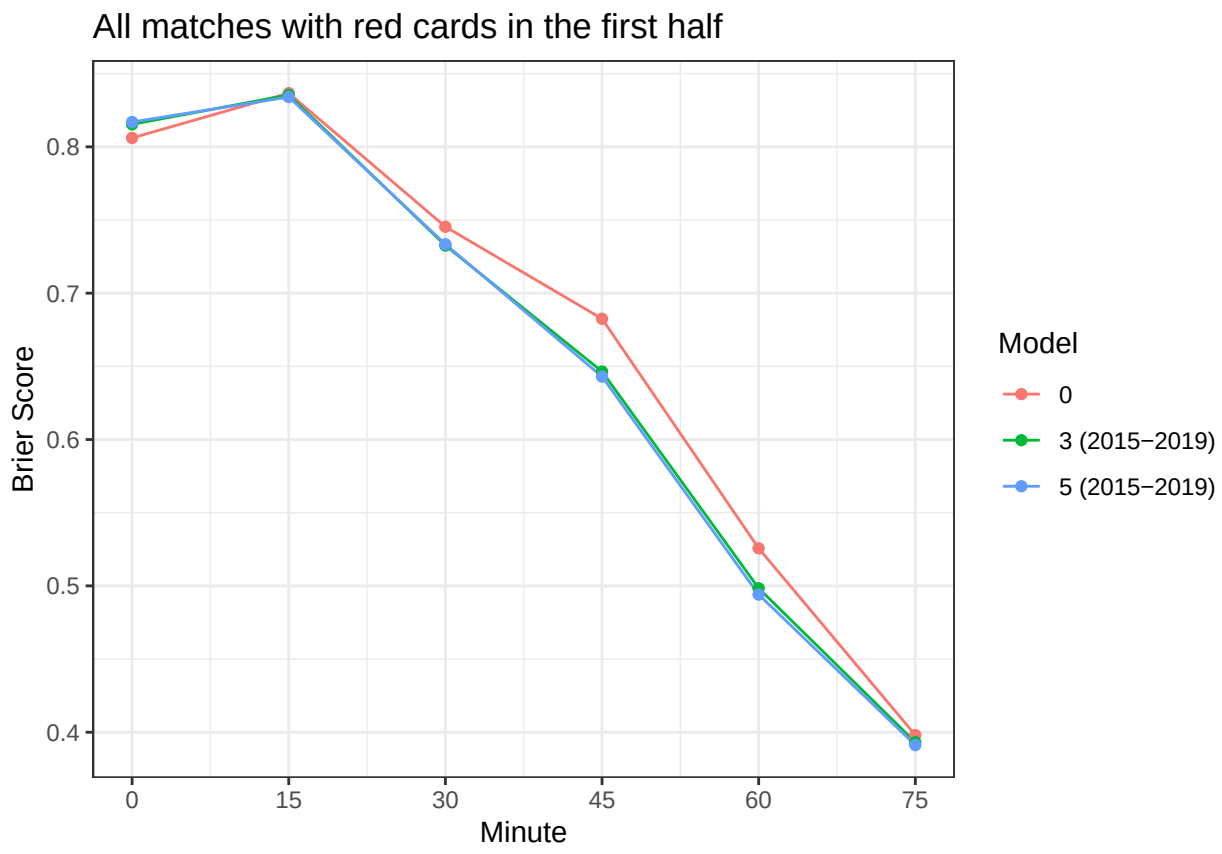
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.6302431	0.5884844	0.5606849	0.5329531	0.4223720	0.3430799
3 (2015-2019)	0.6228720	0.5843654	0.5488402	0.5188887	0.4140975	0.3324337
5 (2015-2019)	0.6230258	0.5847853	0.5478892	0.5186324	0.4142637	0.3316354

```
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	0.8060611	0.8366821	0.7453939	0.6825199	0.5256499	0.3980519
3 (2015-2019)	0.8153712	0.8354634	0.7325725	0.6464705	0.4983613	0.3931800
5 (2015-2019)	0.8169697	0.8340221	0.7334338	0.6429818	0.4940609	0.3911999