Ranked Probability Score

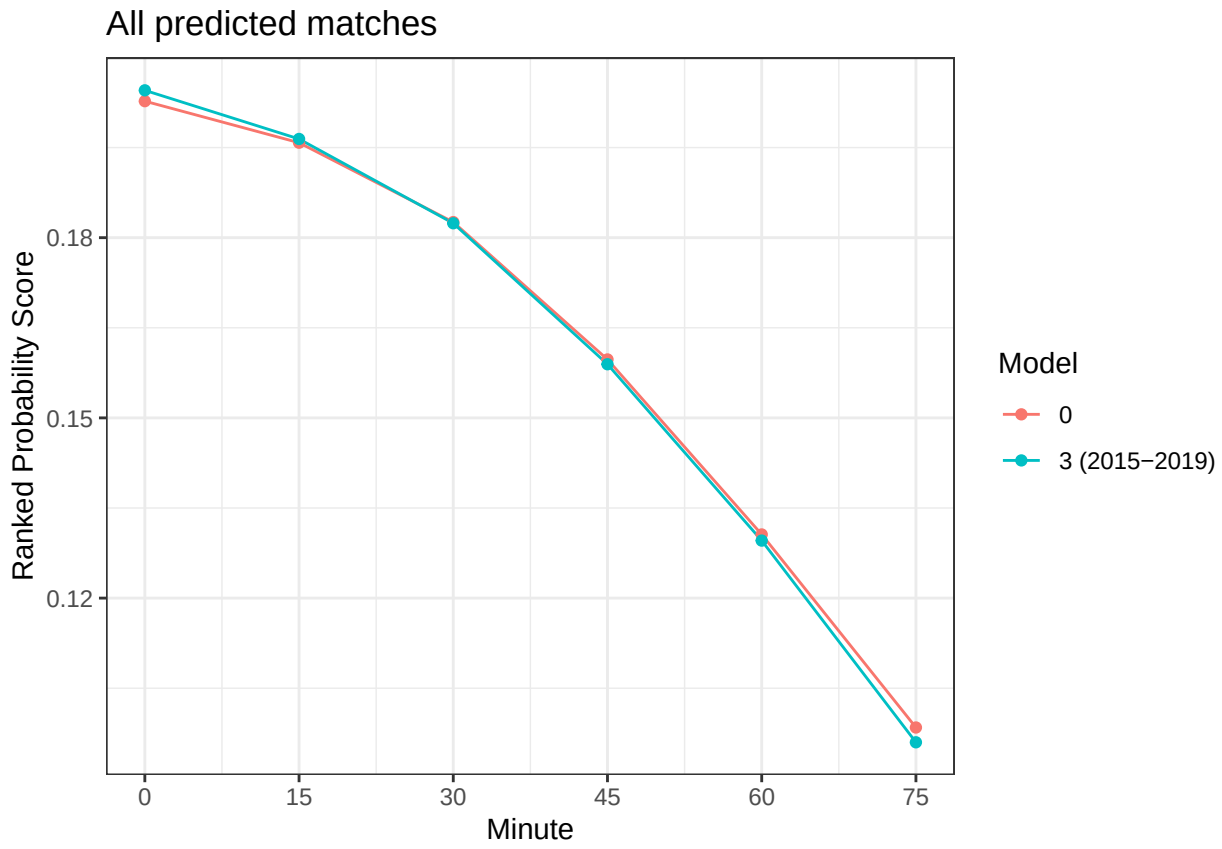
```
library(dplyr)
library(ggplot2)
library(tidyr)
library(knitr)

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

nrow(HDA)
```

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

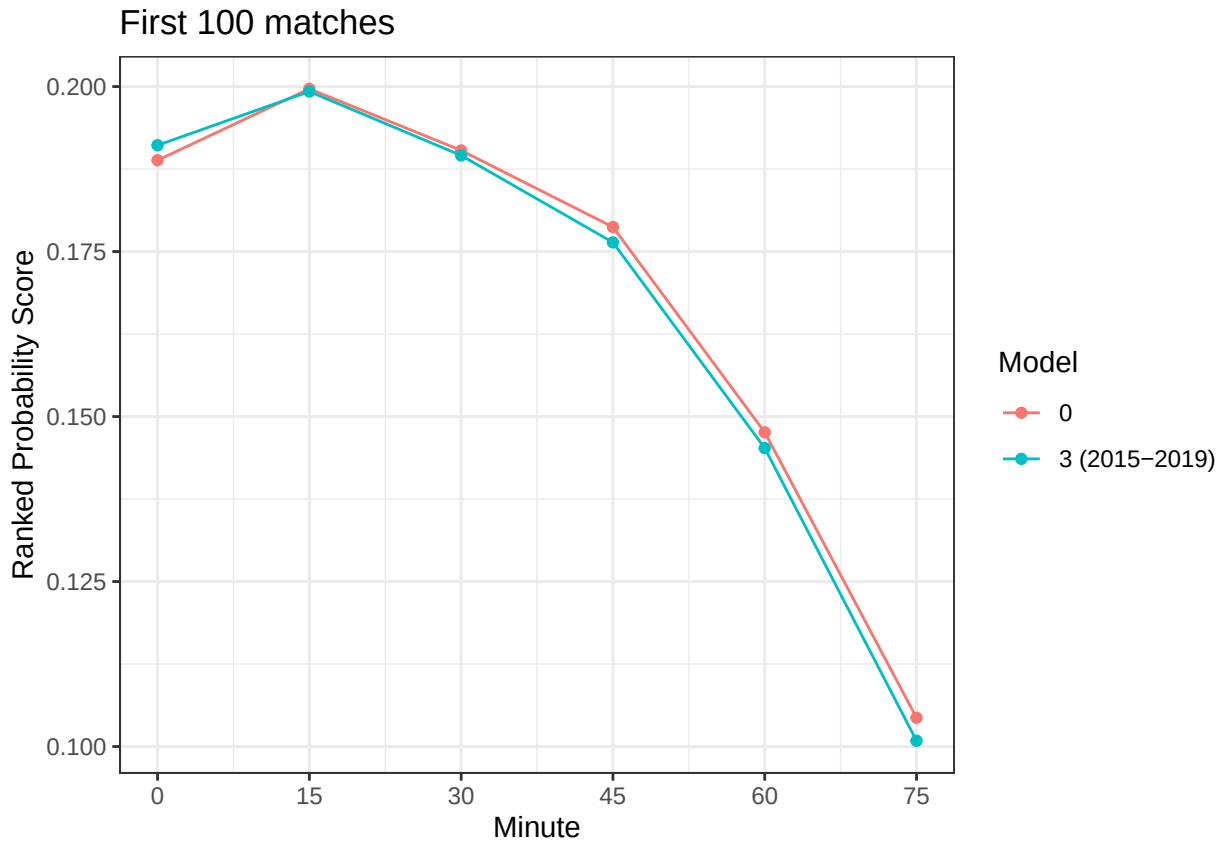
```
all = tibble(RPS = apply(HDA[,c(45:56)], 2, mean),
             Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
             Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                            levels = c("0", "3 (2015-2019)")))
all %>%
  ggplot(aes(x = Minute, y = RPS, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches") +
  ylab("Ranked Probability Score")
```




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

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.2027273	0.1958097	0.1825967	0.1597285	0.1305939	0.0984595
3 (2015-2019)	0.2045287	0.1964316	0.1823990	0.1589316	0.1295744	0.0959938

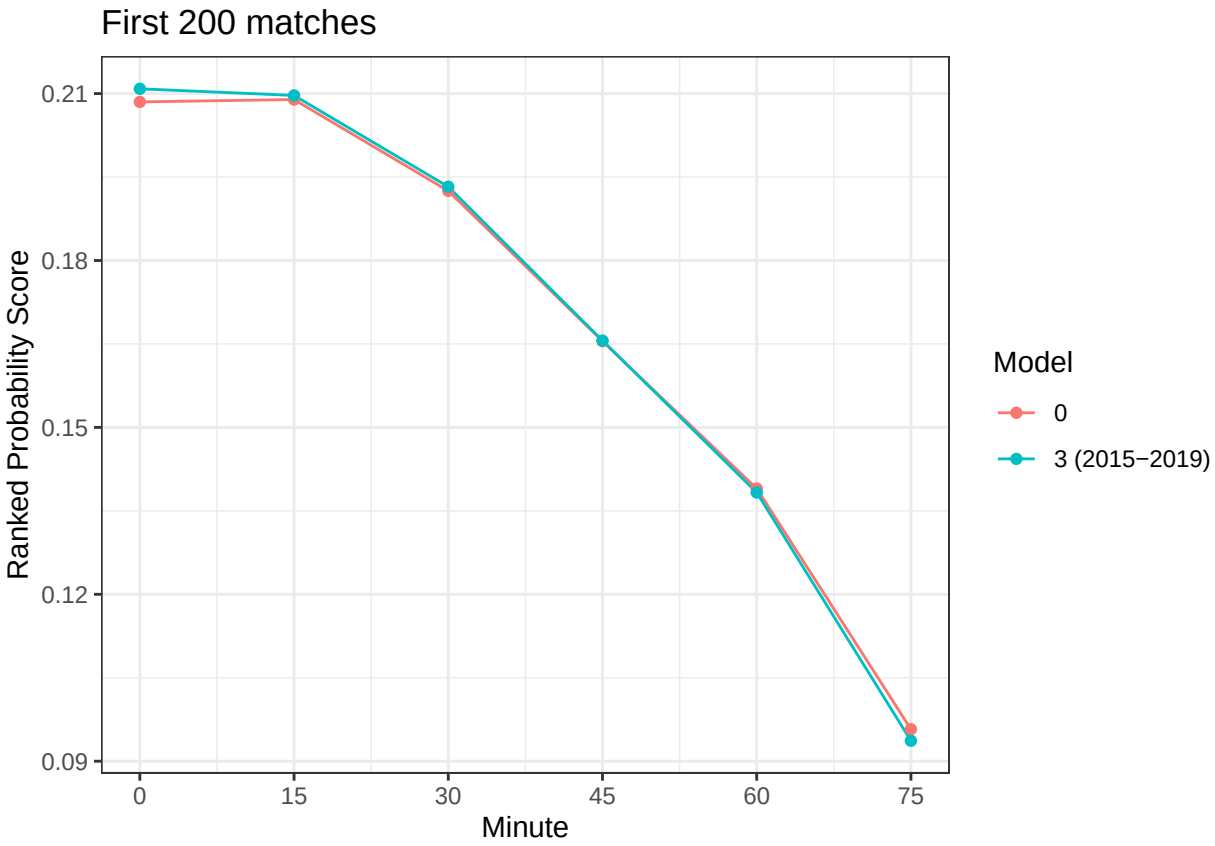
```
first_100 = tibble(RPS = apply(HDA[c(1:100), c(45:56)], 2, mean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                   Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                                  levels = c("0", "3 (2015-2019)")))
first_100 %>%
  ggplot(aes(x = Minute, y = RPS, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("First 100 matches") +
  ylab("Ranked Probability Score")
```




```
first_100 %>%
  pivot_wider(id_cols = "Model", values_from = "RPS", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.1888320	0.1996614	0.1903227	0.1787155	0.1476119	0.1043612
3 (2015-2019)	0.1911062	0.1992424	0.1895588	0.1763870	0.1452201	0.1008752

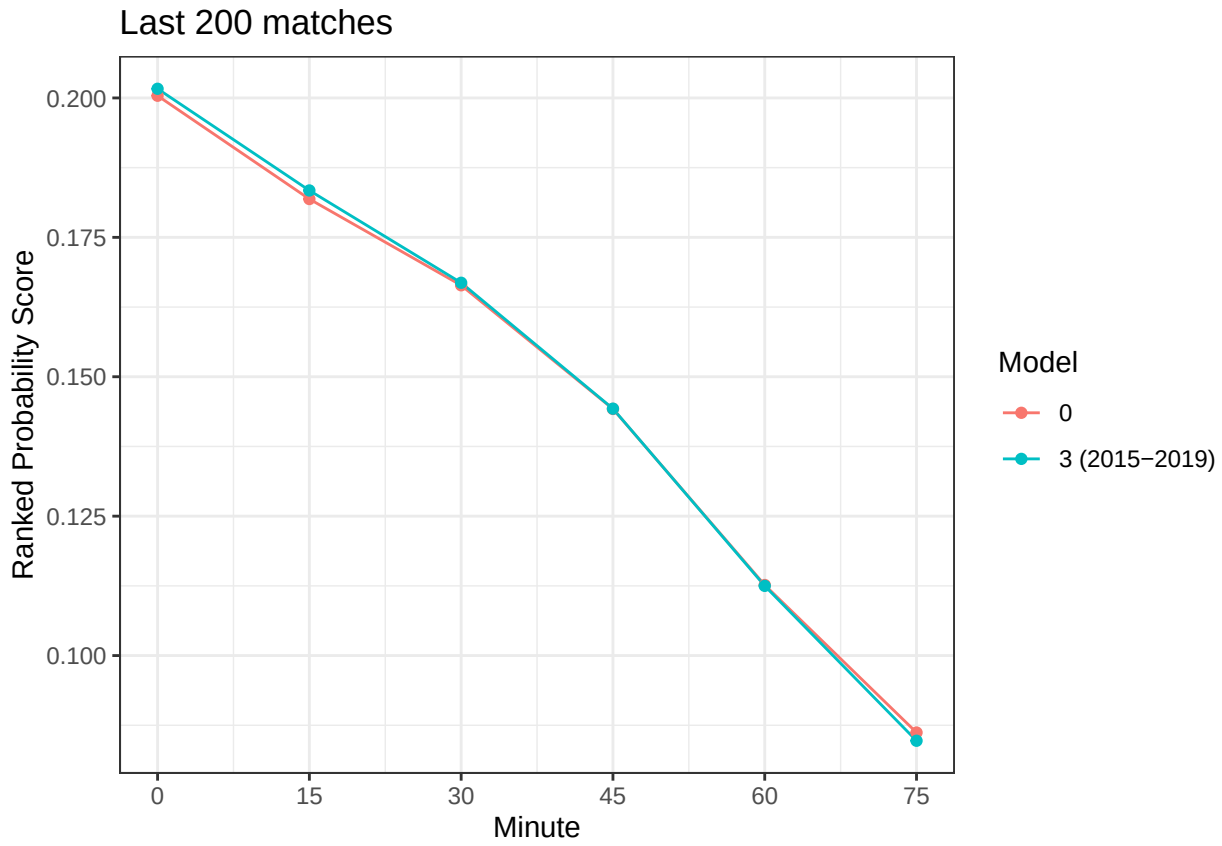
```
first_200 = tibble(RPS = apply(HDA[c(1:200), c(45:56)], 2, mean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                   Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                                  levels = c("0", "3 (2015-2019)")))
first_200 %>%
  ggplot(aes(x = Minute, y = RPS, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("First 200 matches") +
  ylab("Ranked Probability Score")
```




```
first_200 %>%
  pivot_wider(id_cols = "Model", values_from = "RPS", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.2085075	0.2089284	0.1925139	0.1654958	0.1390102	0.0957476
3 (2015-2019)	0.2108631	0.2096706	0.1932620	0.1656019	0.1382953	0.0936729

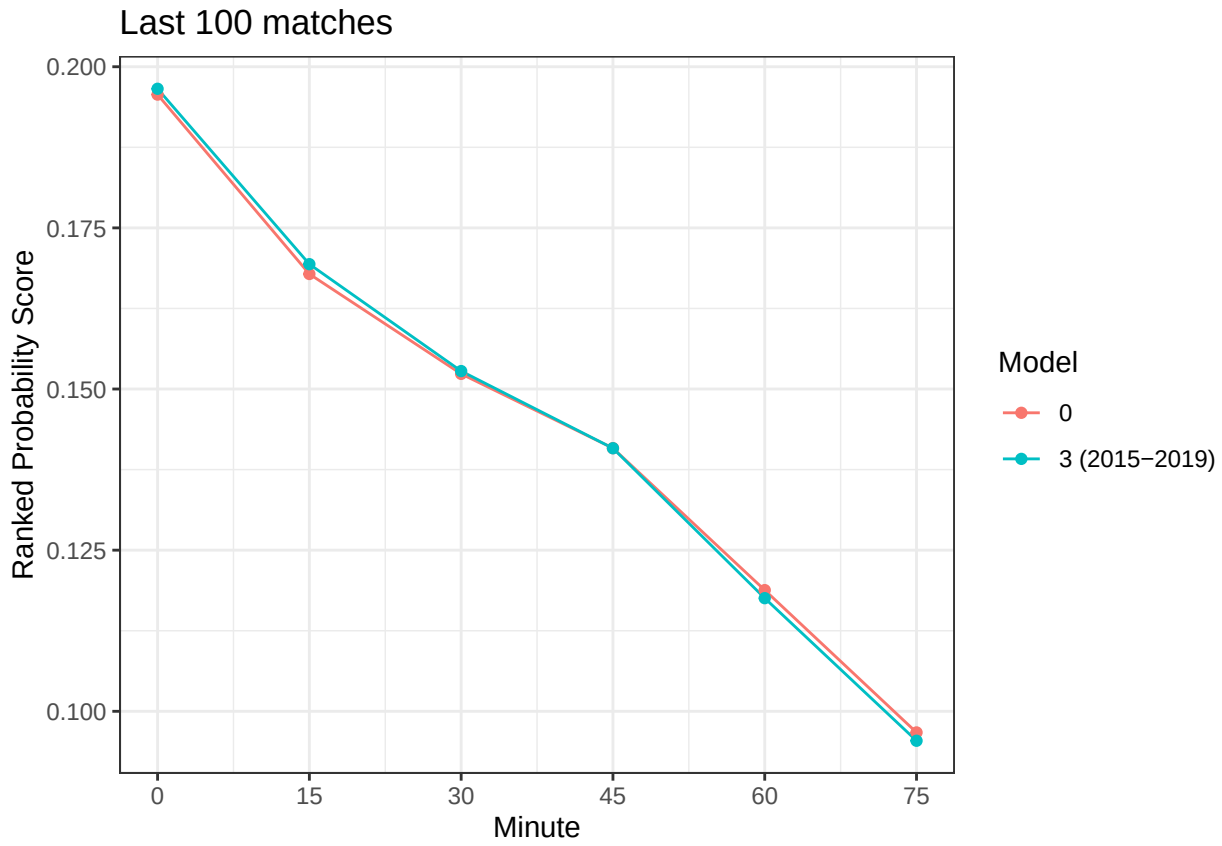
```
last_200 = tibble(RPS = apply(HDA[c(141:340), c(45:56)], 2, mean),
                  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                  Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                                 levels = c("0", "3 (2015-2019)")))
last_200 %>%
  ggplot(aes(x = Minute, y = RPS, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("Last 200 matches") +
  ylab("Ranked Probability Score")
```




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

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.2003968	0.1818714	0.1664027	0.1442278	0.1126557	0.0861980
3 (2015-2019)	0.2016417	0.1834096	0.1668526	0.1442762	0.1124965	0.0847231

```
last_100 = tibble(RPS = apply(HDA[c(241:340), c(45:56)], 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 = RPS, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("Last 100 matches") +
  ylab("Ranked Probability Score")
```




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

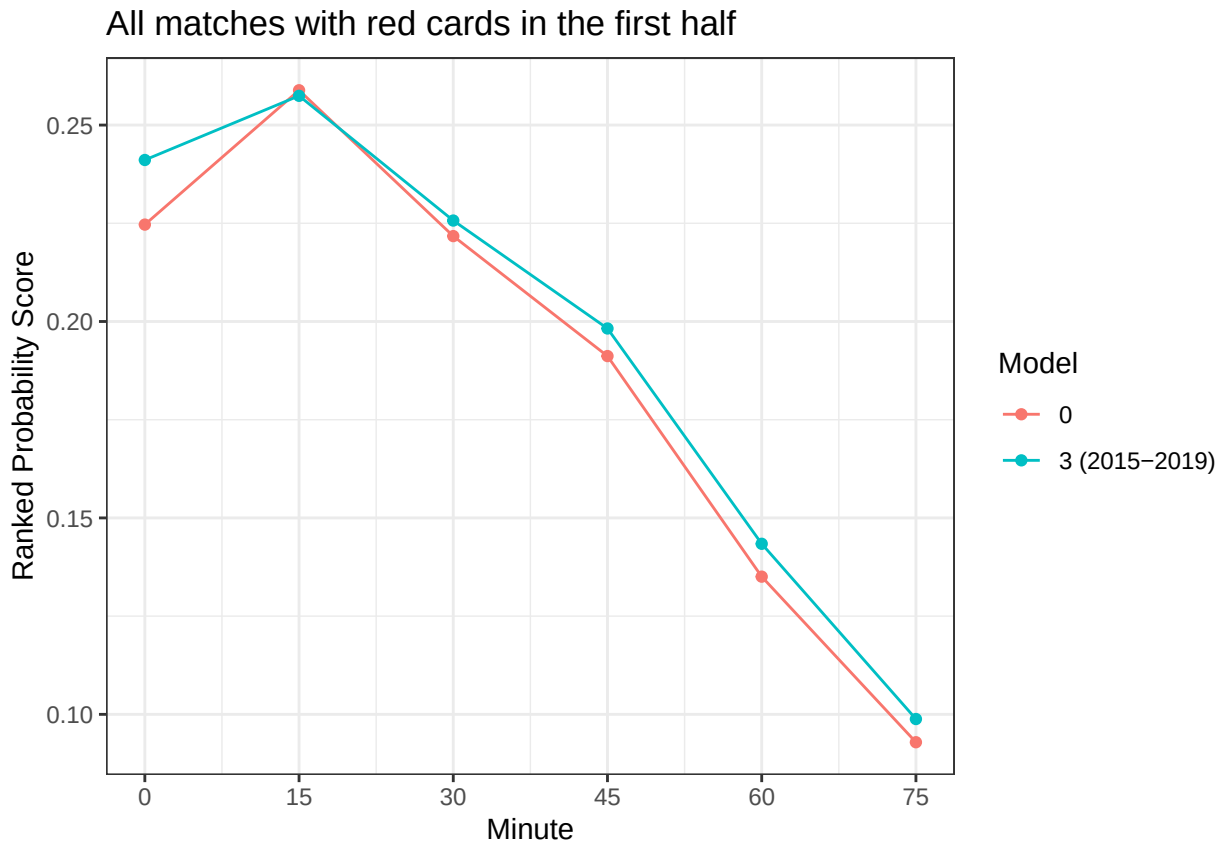
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.1956777	0.1678463	0.1523482	0.1408272	0.1188033	0.0967206
3 (2015-2019)	0.1965787	0.1693699	0.1527839	0.1408105	0.1175529	0.0954313

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

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

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

all_reds = tibble(RPS = apply(HDA_reds[,c(45:56)], 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 = RPS, 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("Ranked Probability Score")
```



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

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.224667	0.2588605	0.221737	0.1911983	0.1350503	0.0929061
3 (2015-2019)	0.241114	0.2574463	0.225699	0.1982264	0.1434174	0.0988127