

Ranked Probability Score

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

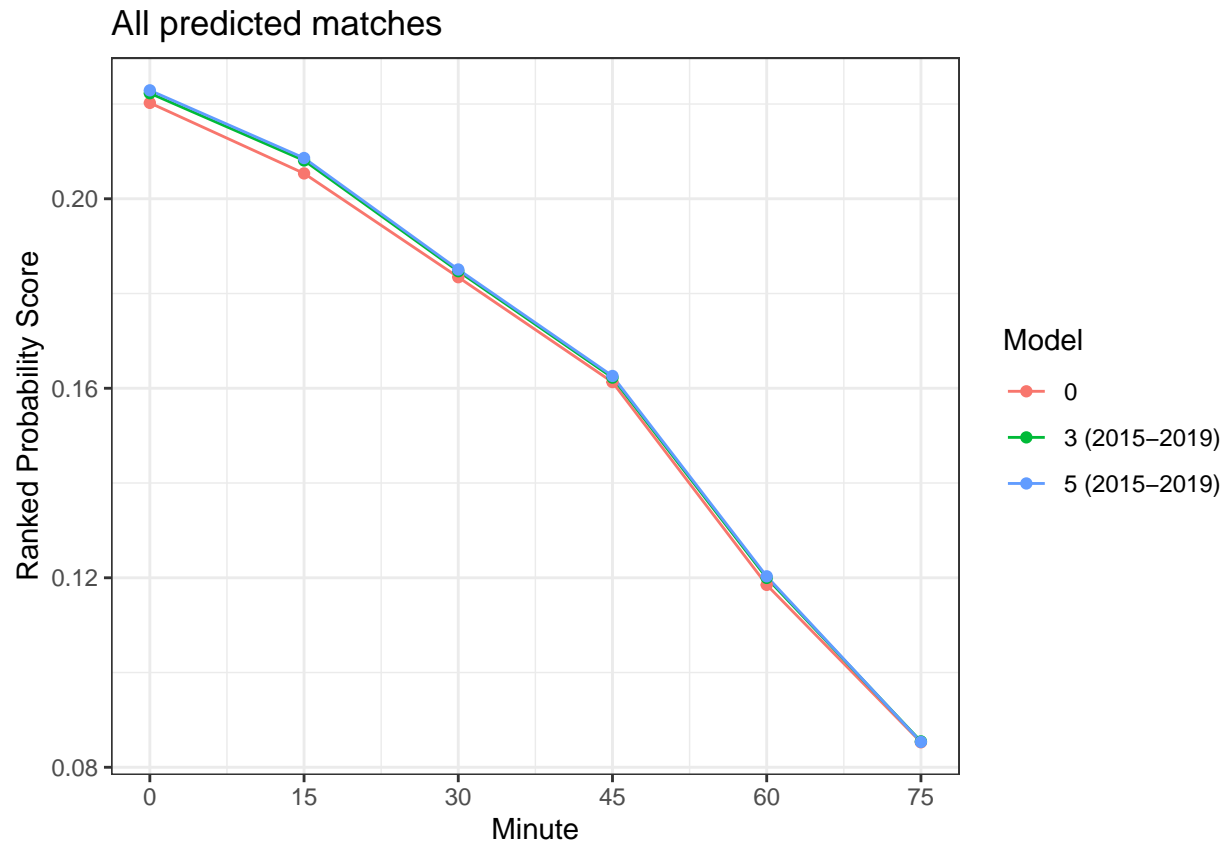
load("data/HDA.RData")
load("~/GitHub/soccer-live-predictions/soccer-live-predictions/scrape/data/reds.RData")

nrow(HDA)
```

```
## [1] 333
```

```
all = tibble(RPS = apply(HDA[,c(99:104, 117:128)], 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 = 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()
```

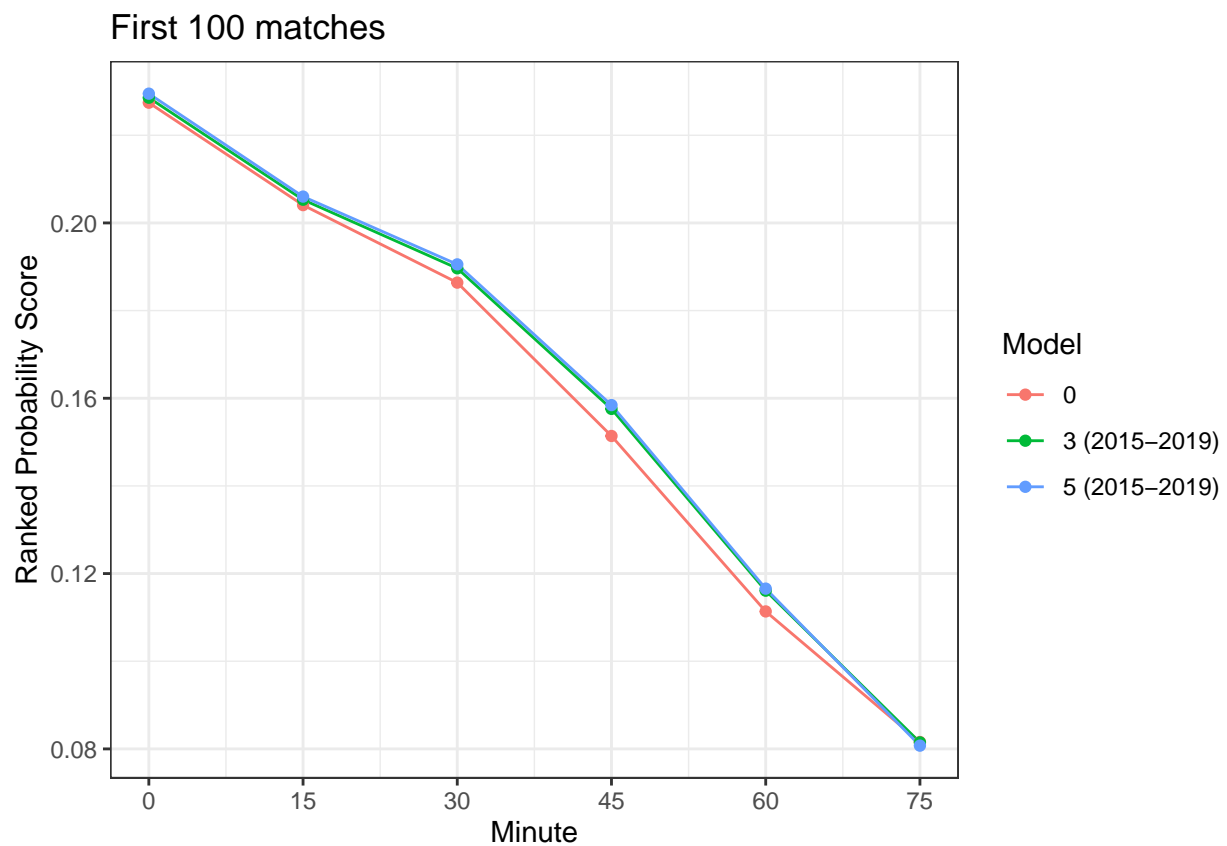
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.2202365	0.2053505	0.1834428	0.1612963	0.1184992	0.0852712
3 (2015-2019)	0.2222785	0.2080854	0.1847107	0.1622544	0.1199706	0.0854637
5 (2015-2019)	0.2228871	0.2085984	0.1850695	0.1625932	0.1203066	0.0853594

```

first_100 = tibble(RPS = apply(HDA[c(1:100), c(99:104, 117:128)], 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 = 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()

```

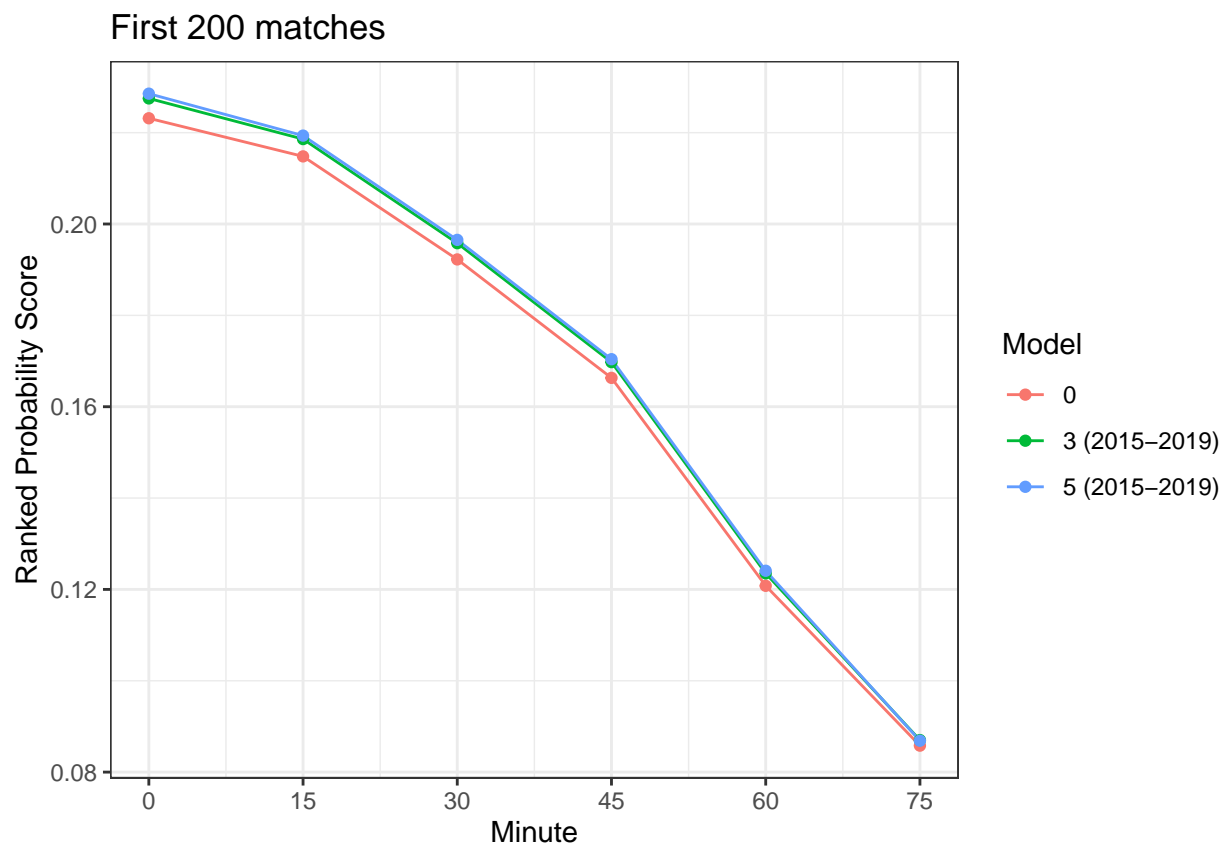
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.2274295	0.2040659	0.1863819	0.1513934	0.1113637	0.0815728
3 (2015-2019)	0.2285686	0.2052946	0.1896470	0.1575728	0.1160949	0.0814845
5 (2015-2019)	0.2294930	0.2060105	0.1905440	0.1584420	0.1165644	0.0807430

```

first_200 = tibble(RPS = apply(HDA[c(1:200)], c(99:104, 117:128)], 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 = 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")

```



```

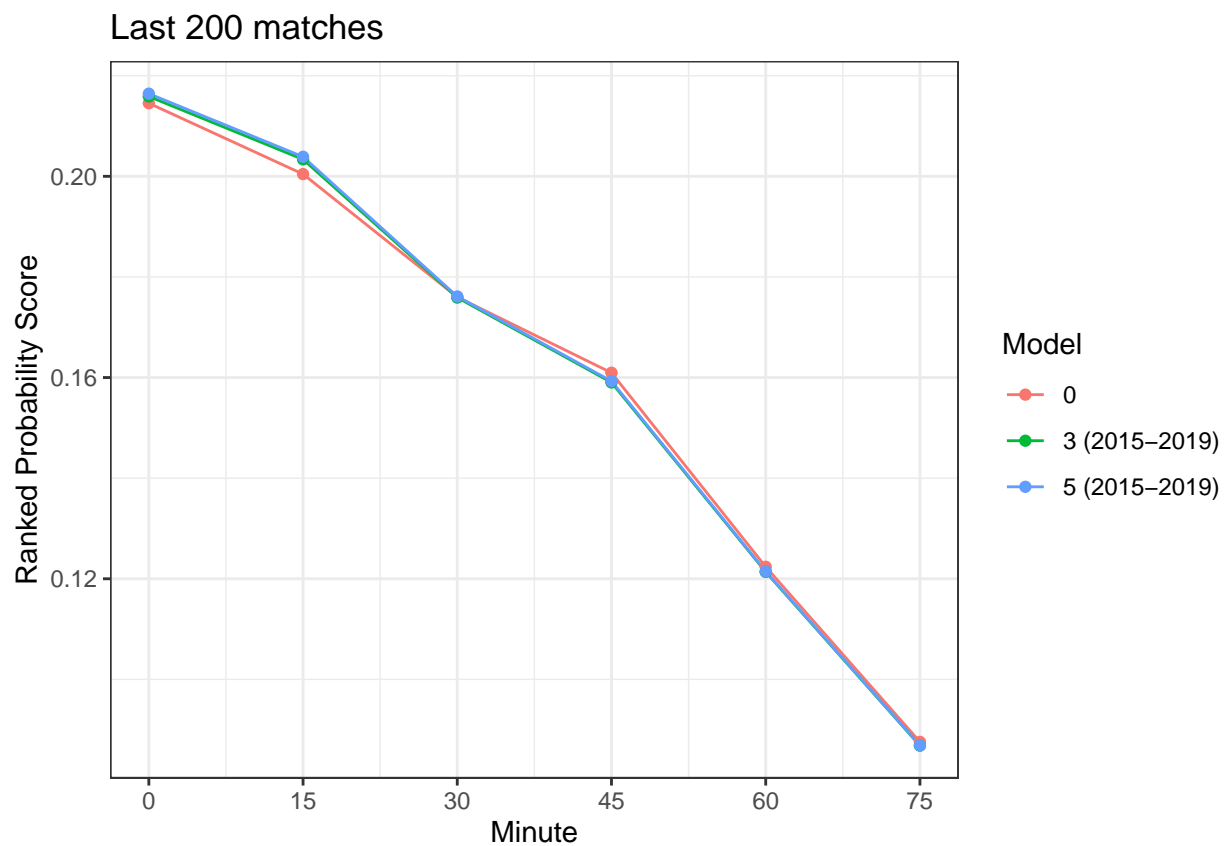
first_200 %>%
  pivot_wider(id_cols = "Model", values_from = "RPS", names_from = "Minute",
    names_prefix = "Minute ") %>%
  kable()

```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.2231531	0.2148029	0.1922461	0.1663102	0.1207875	0.0857964
3 (2015-2019)	0.2274963	0.2185911	0.1958012	0.1697543	0.1235113	0.0870163
5 (2015-2019)	0.2285349	0.2193499	0.1965248	0.1704042	0.1240836	0.0868773

```
last_200 = tibble(RPS = apply(HDA[c(134:333), c(99:104, 117:128)], 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 = 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")
```

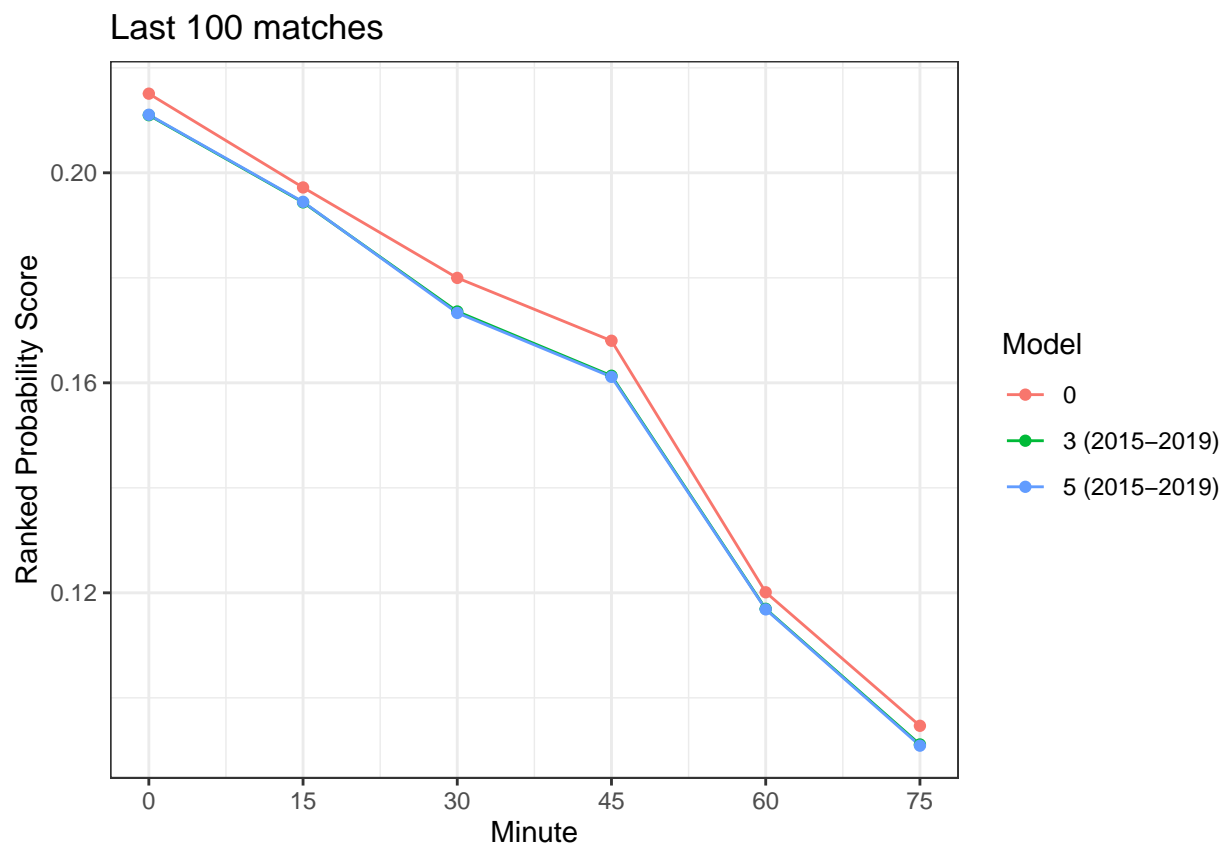


```
last_200 %>%
  pivot_wider(id_cols = "Model", values_from = "RPS", names_from = "Minute",
    names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.2145304	0.2004427	0.1760006	0.1609501	0.1223754	0.0875738
3 (2015-2019)	0.2158734	0.2033686	0.1759044	0.1590262	0.1213591	0.0868270
5 (2015-2019)	0.2164315	0.2038869	0.1761163	0.1592799	0.1214593	0.0868864

```
last_100 = tibble(RPS = apply(HDA[c(234:333), c(99:104, 117:128)], 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 = 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	0.2150731	0.1972086	0.1799835	0.1680088	0.1201060	0.0946721
3 (2015-2019)	0.2109743	0.1943738	0.1735835	0.1613448	0.1169494	0.0911196
5 (2015-2019)	0.2110563	0.1944641	0.1733110	0.1611309	0.1168496	0.0908937

```

matches = reds %>%
  filter(Season == 2020, Half == 1) %>%
  .$Match
length(matches)

```

```
## [1] 23
```

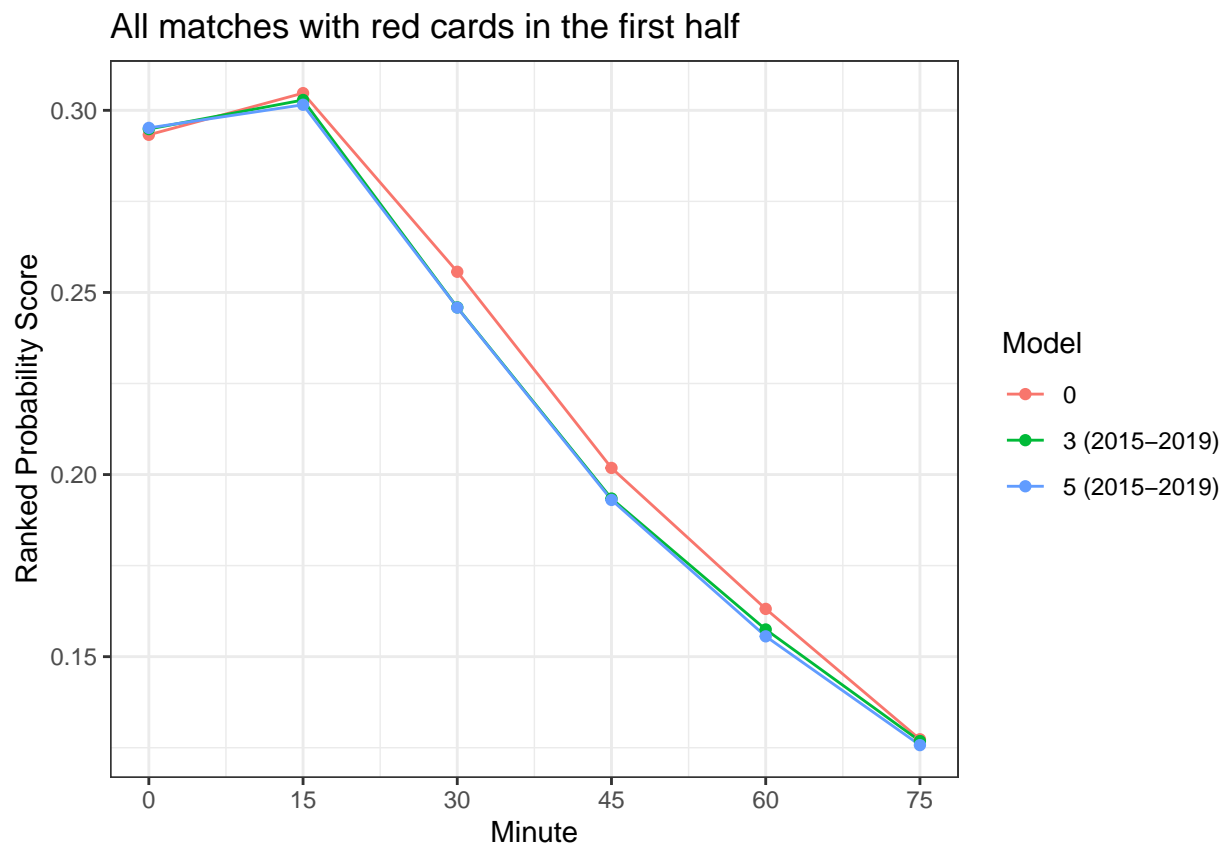
```

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

all_reds = tibble(RPS = apply(HDA_reds[,c(99:104, 117:128)], 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 = 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_recs %>%
  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	0.2932780	0.3047198	0.2556753	0.2018152	0.1631179	0.1273179
3 (2015-2019)	0.2948601	0.3028371	0.2459006	0.1934150	0.1574659	0.1268314
5 (2015-2019)	0.2951766	0.3015266	0.2458420	0.1930328	0.1555842	0.1257170