

## Ranked Probability Score

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

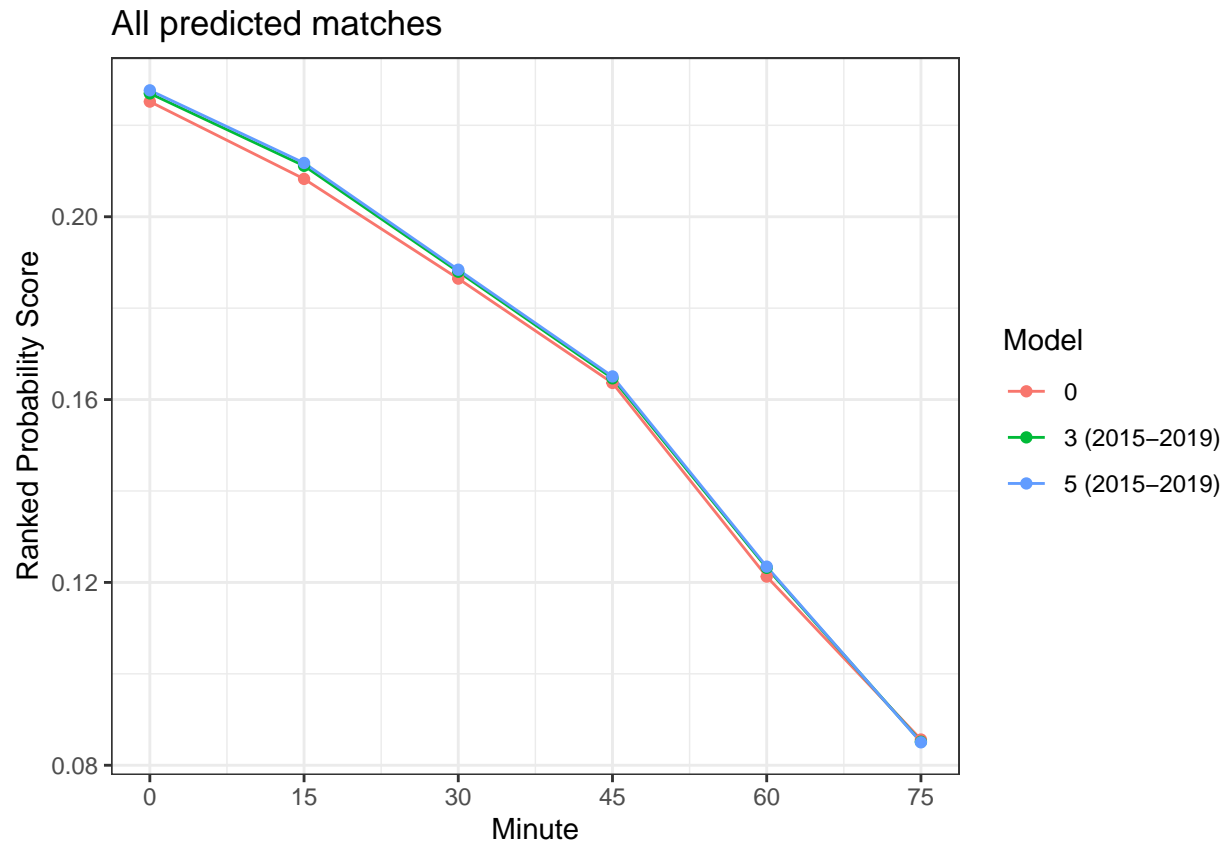
load("data/HDA.RData")
load("~/GitHub/soccer-live-predictions/soccer-live-predictions/scrape/data/reds.RData")

nrow(HDA)
```

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

```
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()
```

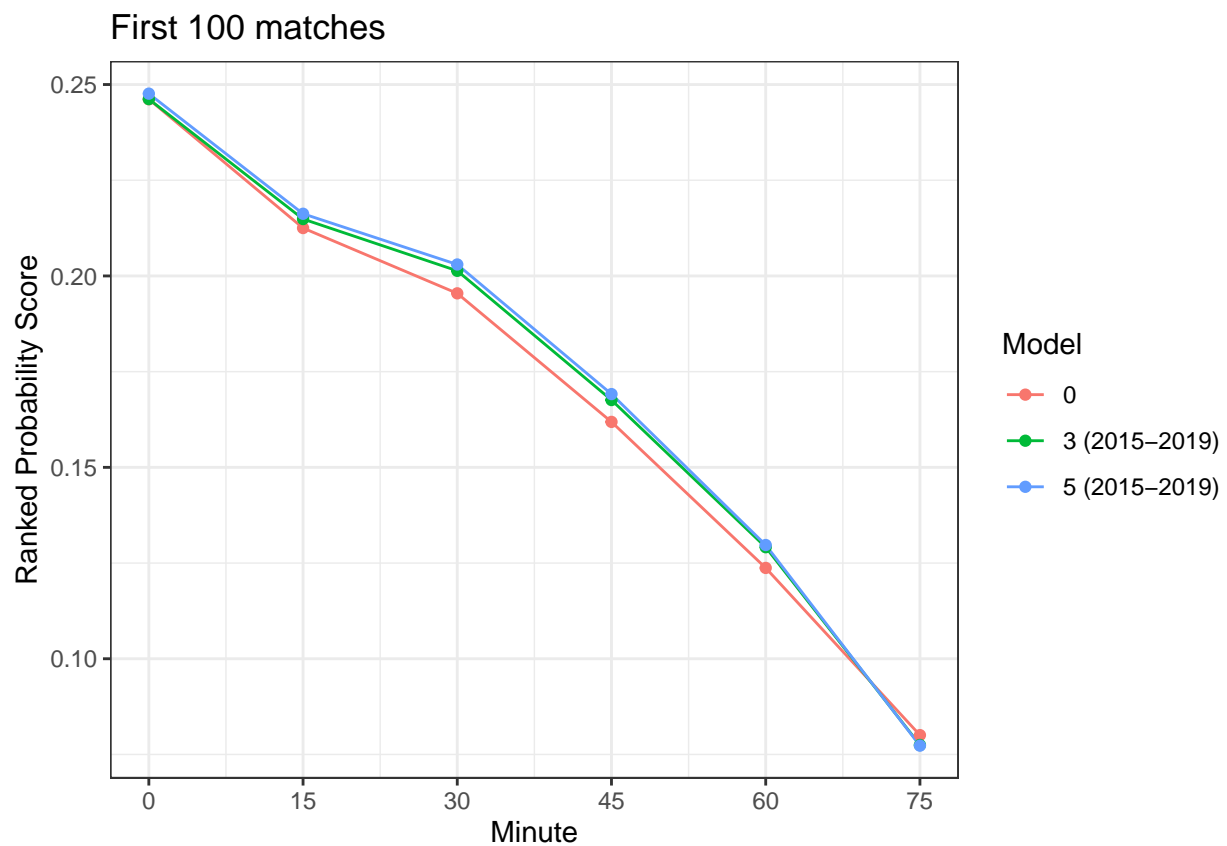
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.2251835	0.2082754	0.1864610	0.1636361	0.1212750	0.0856290
3 (2015-2019)	0.2269793	0.2111423	0.1879707	0.1646484	0.1231845	0.0851382
5 (2015-2019)	0.2276514	0.2117604	0.1884058	0.1650617	0.1234481	0.0850202

```

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()

```

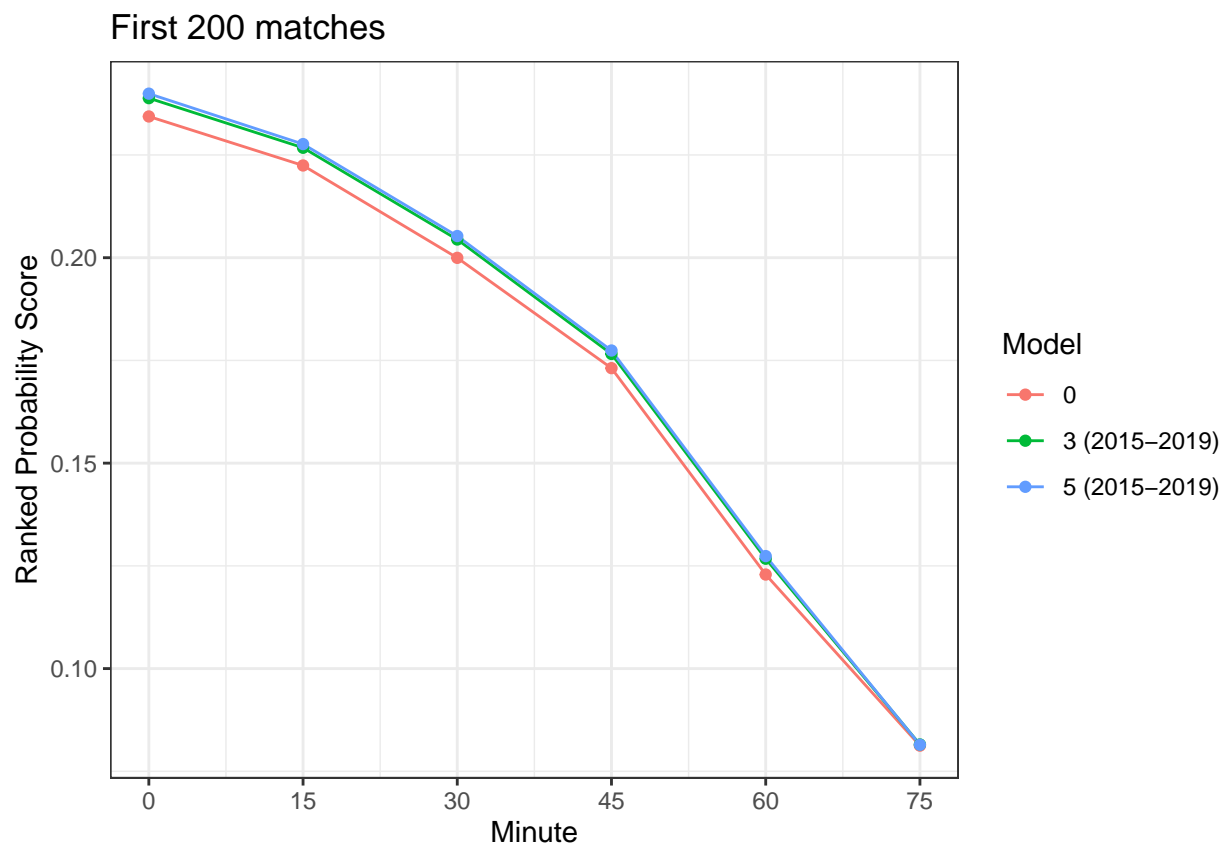
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.2461685	0.2125151	0.1954546	0.1618741	0.1237257	0.0800527
3 (2015-2019)	0.2461946	0.2148664	0.2013440	0.1675877	0.1291368	0.0774757
5 (2015-2019)	0.2476160	0.2162439	0.2029905	0.1691398	0.1297128	0.0773046

```

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()

```

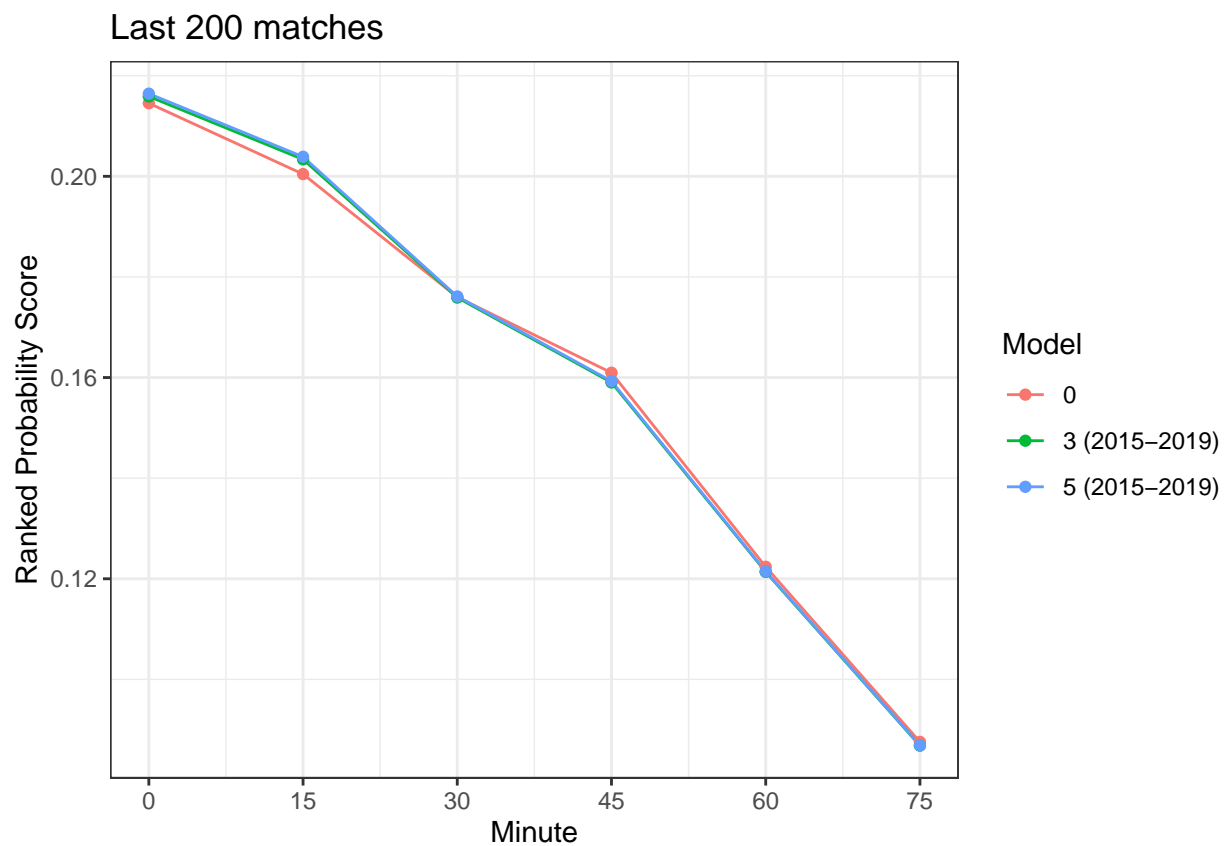
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.2343722	0.2224422	0.1999634	0.1731189	0.1228871	0.0812487
3 (2015-2019)	0.2388319	0.2267541	0.2044263	0.1765747	0.1267333	0.0815321
5 (2015-2019)	0.2399183	0.2276450	0.2052727	0.1774069	0.1273810	0.0814673

```

last_200 = tibble(RPS = apply(HDA[c(151:350)], 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")

```



```

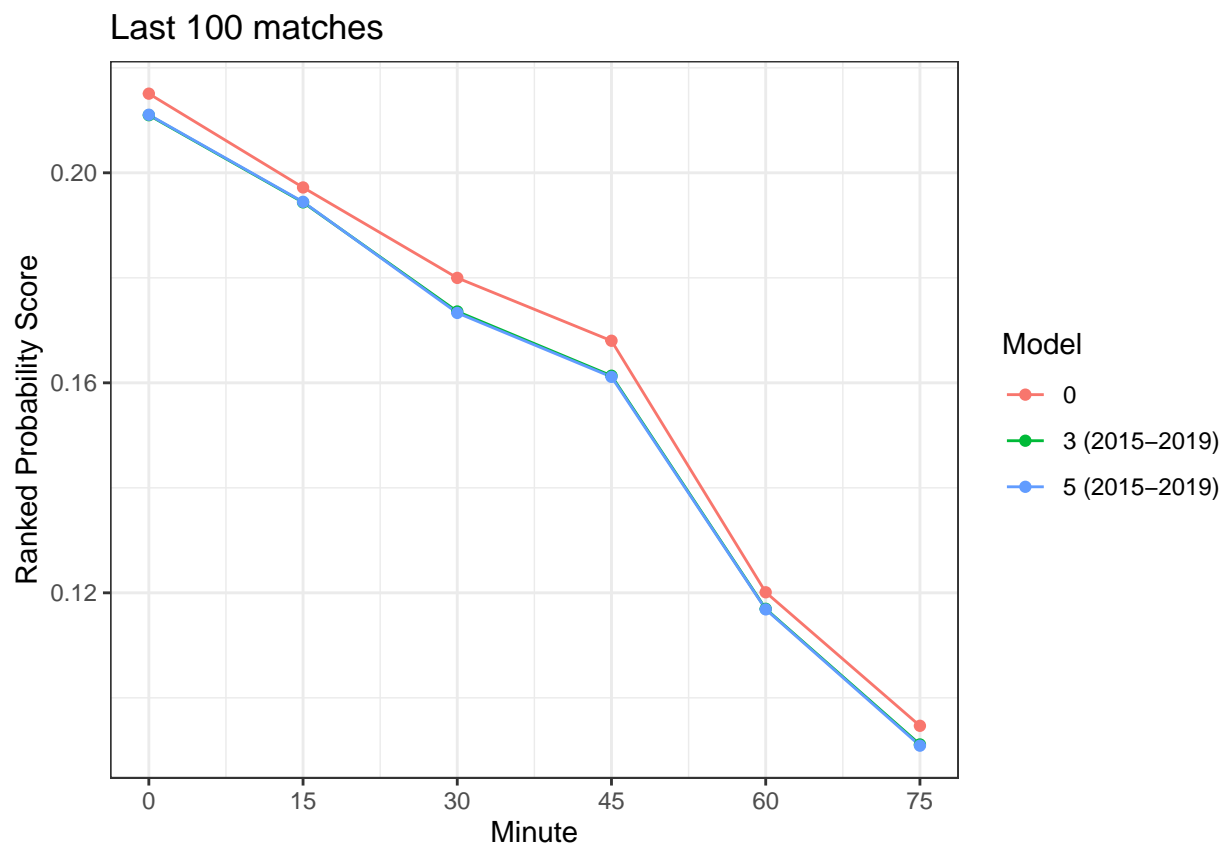
last_200 %>%
  pivot_wider(id_cols = "Model", values_from = "RPS", names_from = "Minute",
    names_prefix = "Minute ") %>%
  kable()

```

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(251:350), 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
```

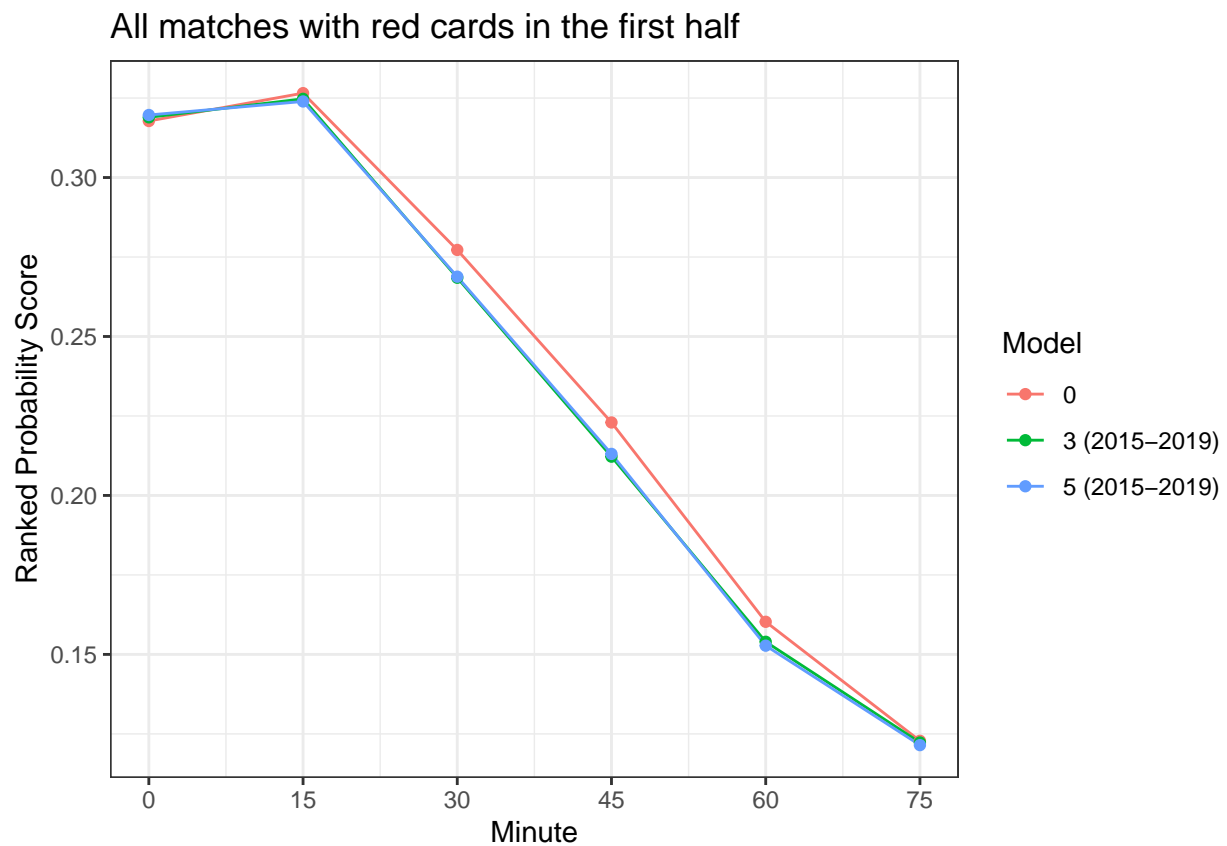
```

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	0.3177855	0.3265452	0.2772292	0.2229595	0.1602613	0.1228243
3 (2015-2019)	0.3189417	0.3247528	0.2684778	0.2122064	0.1539697	0.1222919
5 (2015-2019)	0.3196665	0.3239270	0.2688166	0.2130644	0.1527546	0.1214730