

## Ranked Probability Score

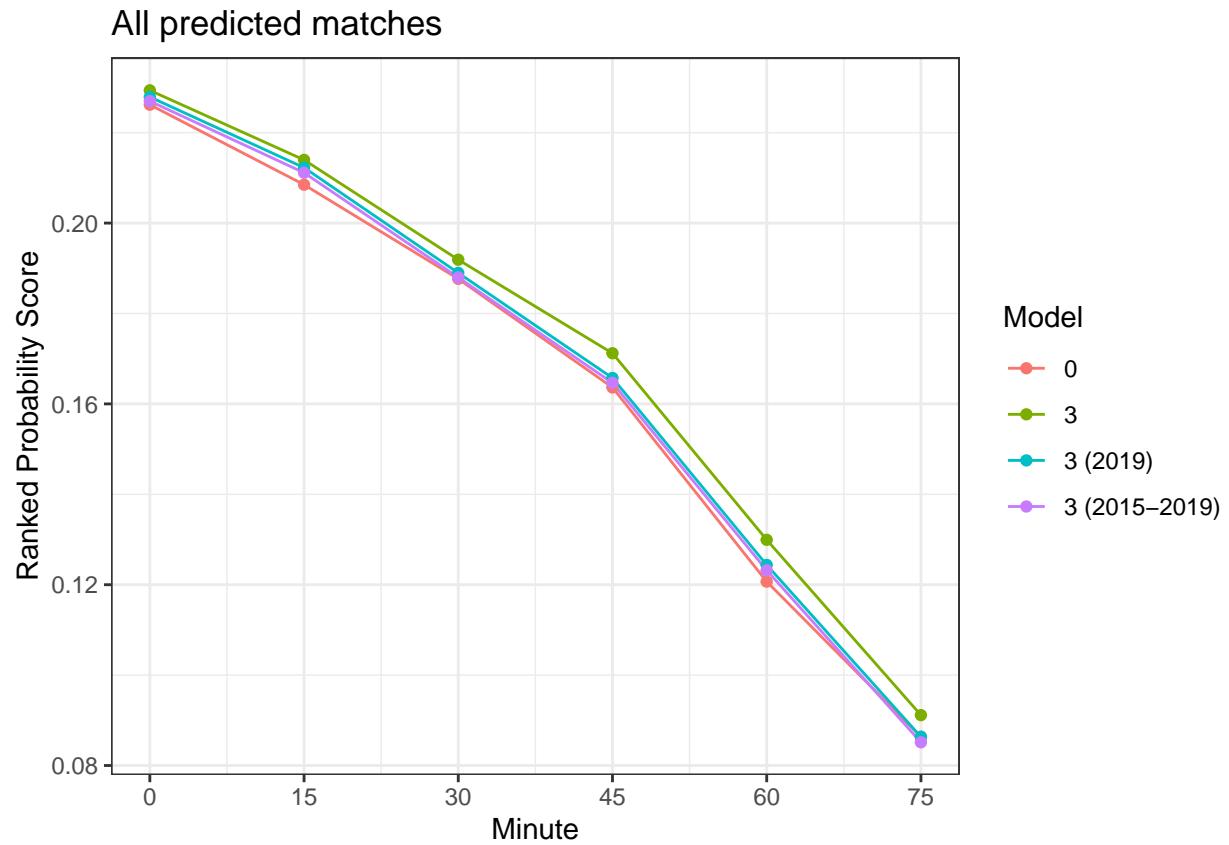
```
library(dplyr)
library(ggplot2)
library(tidyr)
library(knitr)

load("data/HDA.RData")

nrow(HDA)

## [1] 350

all = tibble(RPS = apply(HDA[,c(81:104)], 2, mean),
             Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 4)),
             Model = factor(c(rep("0", 6), rep("3", 6), rep("3 (2019)", 6),
                             rep("3 (2015-2019)", 6)),
                           levels = c("0", "3", "3 (2019)", "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()
```

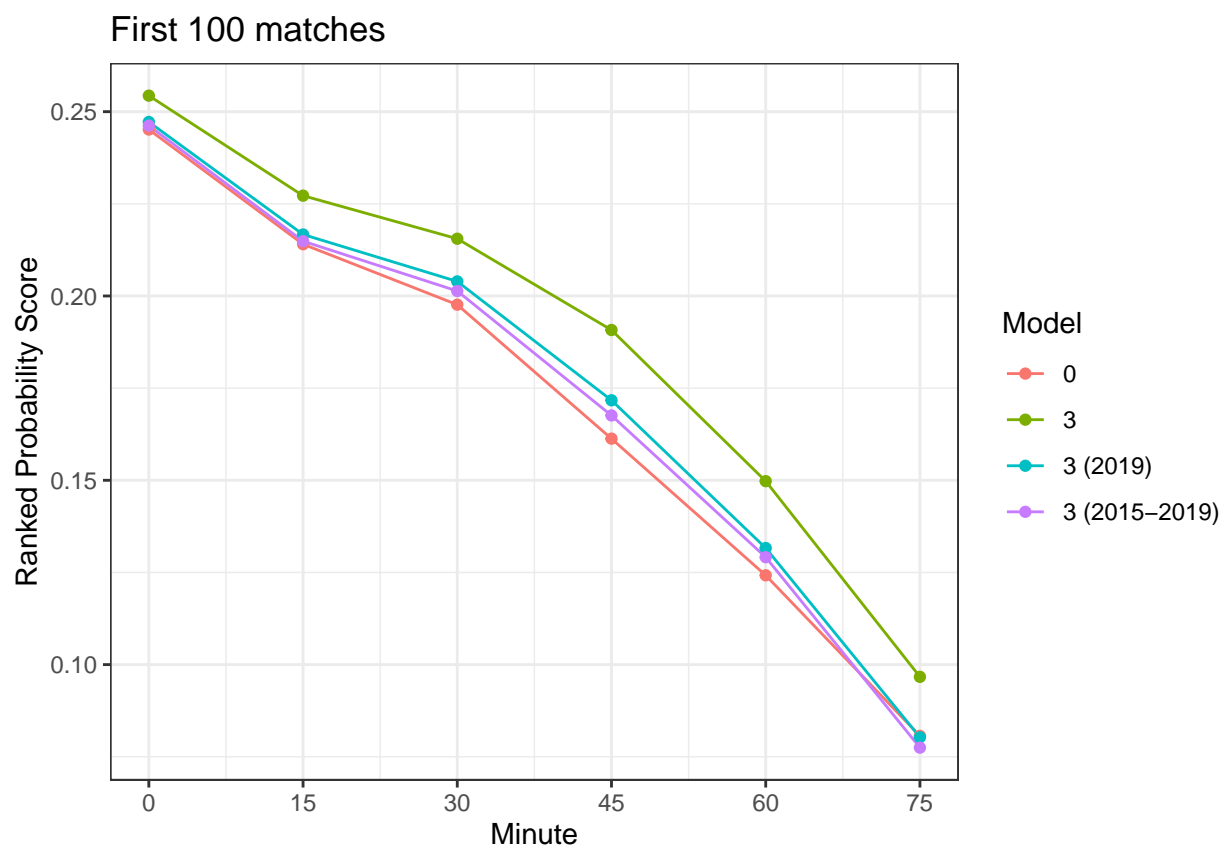
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.2261857	0.2084903	0.1876652	0.1636376	0.1206817	0.0863650
3	0.2293659	0.2139781	0.1918965	0.1712051	0.1299292	0.0911476
3 (2019)	0.2279030	0.2122651	0.1889786	0.1657339	0.1243726	0.0863460
3 (2015-2019)	0.2269793	0.2111423	0.1879707	0.1646484	0.1231845	0.0851382

```

first_100 = tibble(RPS = apply(HDA[c(1:100),c(81:104)], 2, mean),
                    Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 4)),
                    Model = factor(c(rep("0", 6), rep("3", 6), rep("3 (2019)", 6),
                                     rep("3 (2015-2019)", 6)),
                                   levels = c("0", "3", "3 (2019)", "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()

```

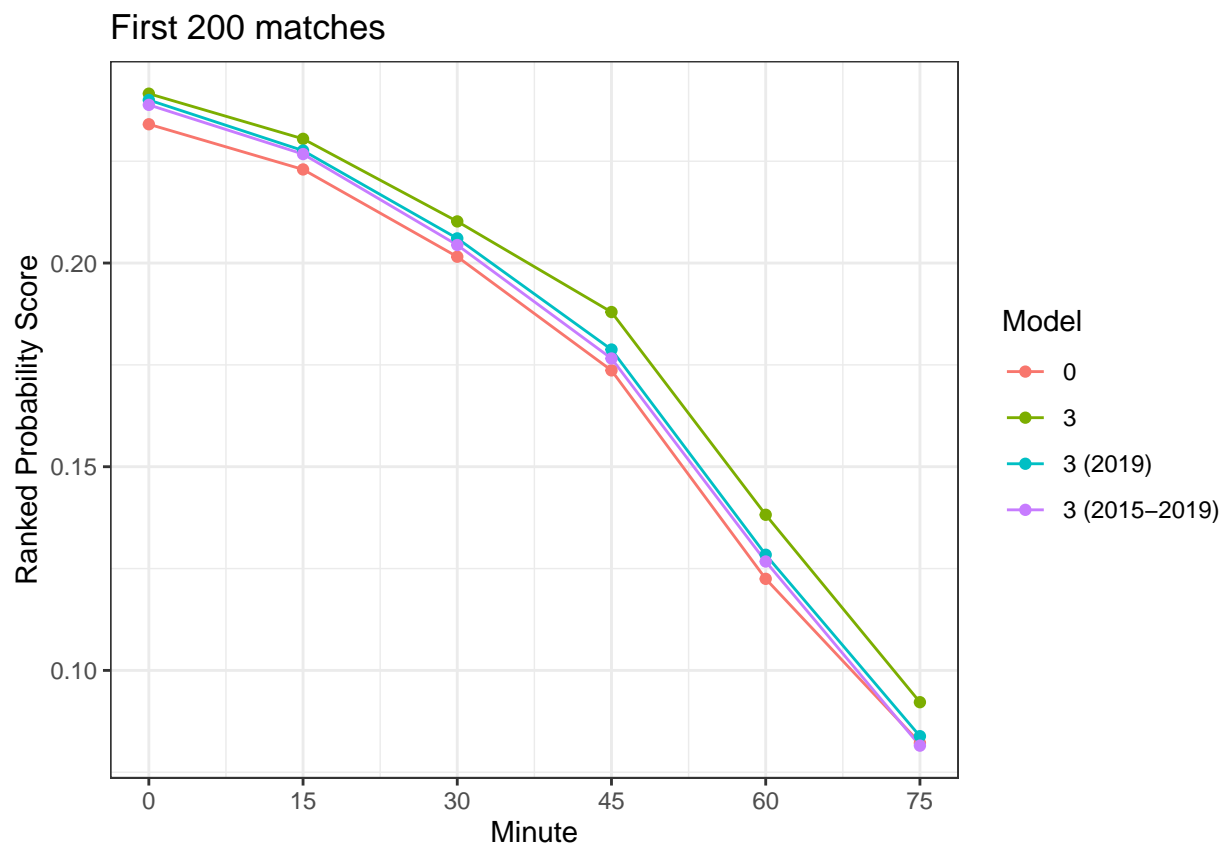
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.2451603	0.2140415	0.1976387	0.1612957	0.1242198	0.0806600
3	0.2543389	0.2272117	0.2155270	0.1907664	0.1497635	0.0966810
3 (2019)	0.2472101	0.2166834	0.2039458	0.1717072	0.1316214	0.0802863
3 (2015-2019)	0.2461946	0.2148664	0.2013440	0.1675877	0.1291368	0.0774757

```

first_200 = tibble(RPS = apply(HDA[c(1:200),c(81:104)], 2, mean),
                    Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 4)),
                    Model = factor(c(rep("0", 6), rep("3", 6), rep("3 (2019)", 6),
                                    rep("3 (2015-2019)", 6)),
                                levels = c("0", "3", "3 (2019)", "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()

```

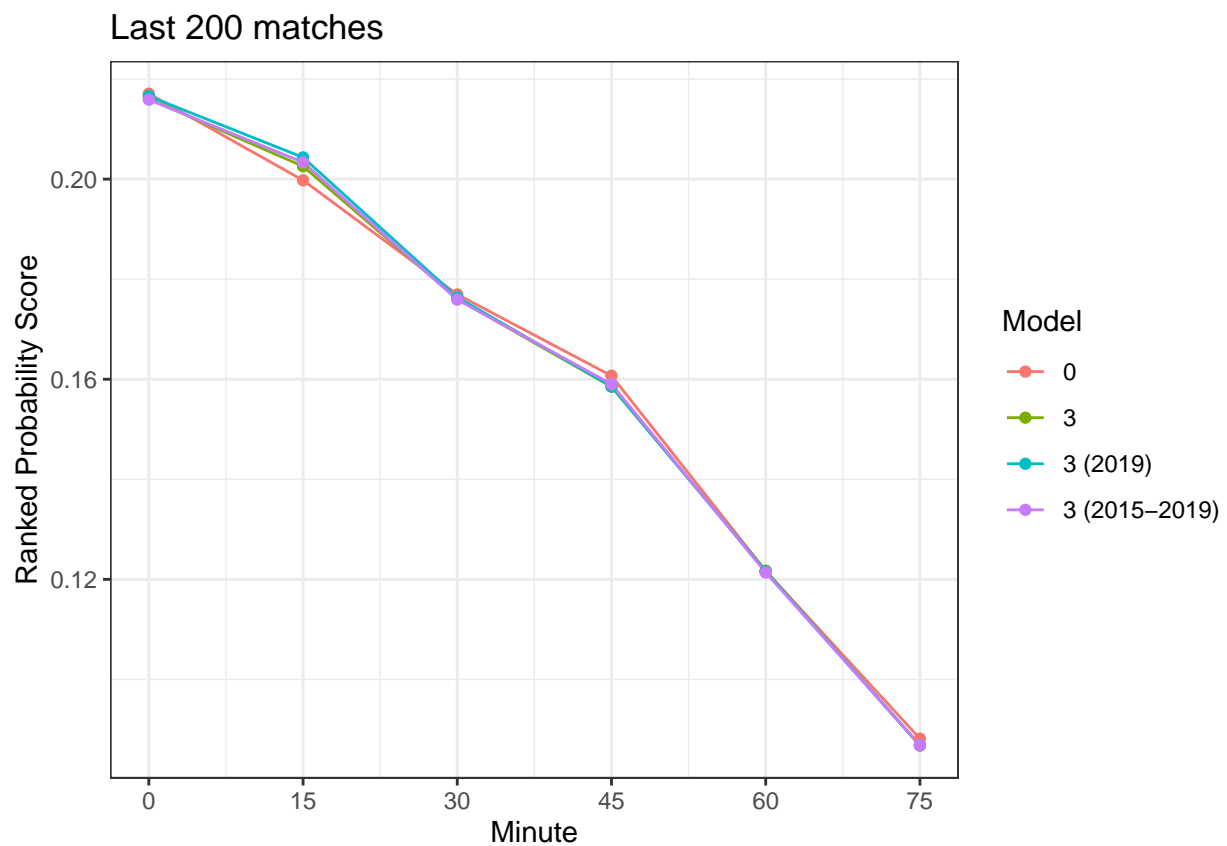
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.2340741	0.2229745	0.2015535	0.1736492	0.1224783	0.0821848
3	0.2415802	0.2304903	0.2102192	0.1879734	0.1382016	0.0921972
3 (2019)	0.2400275	0.2276091	0.2060563	0.1787687	0.1283917	0.0838548
3 (2015-2019)	0.2388319	0.2267541	0.2044263	0.1765747	0.1267333	0.0815321

```

last_200 = tibble(RPS = apply(HDA[c(151:350),c(81:104)], 2, mean),
  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 4)),
  Model = factor(c(rep("0", 6), rep("3", 6), rep("3 (2019)", 6),
    rep("3 (2015-2019)", 6)),
    levels = c("0", "3", "3 (2019)", "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()

```

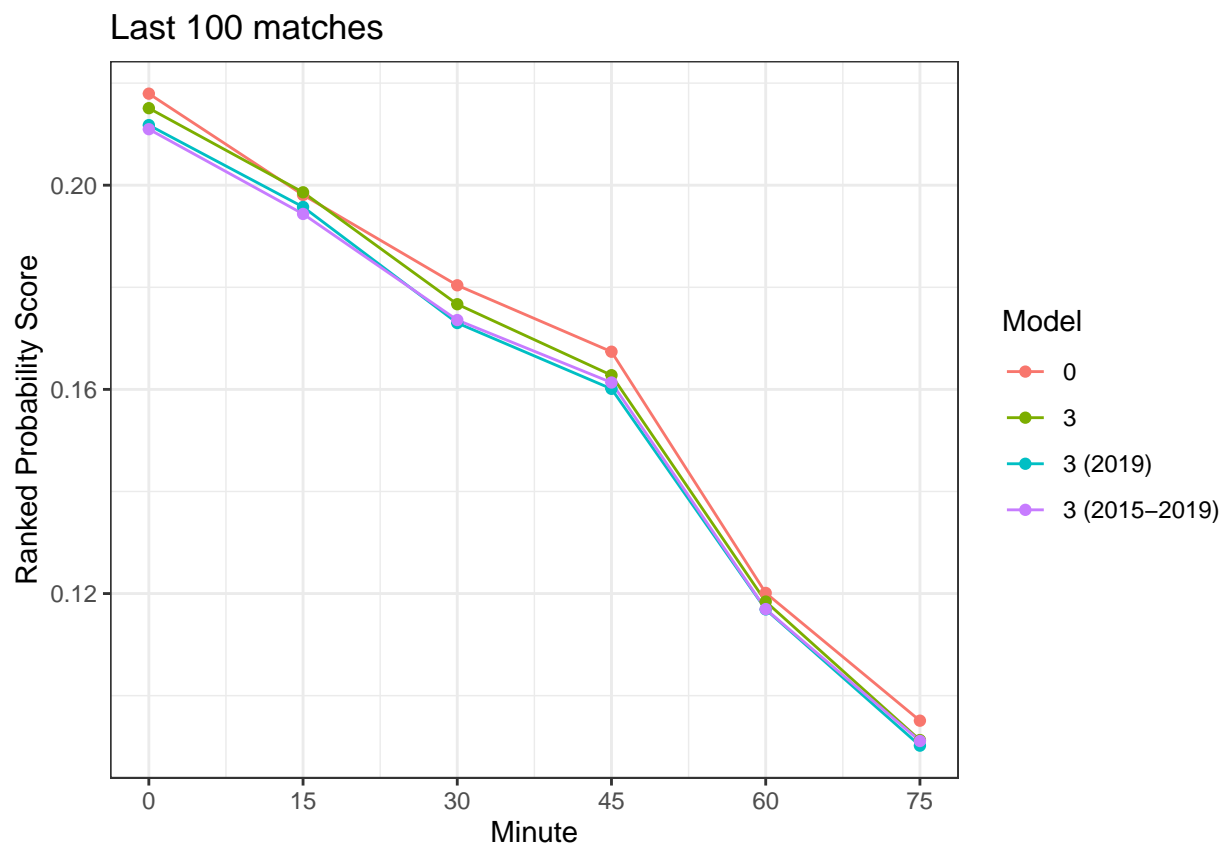
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.2170793	0.1997681	0.1769500	0.1606936	0.1216809	0.0881521
3	0.2160553	0.2025629	0.1760475	0.1584983	0.1217008	0.0867588
3 (2019)	0.2165927	0.2043409	0.1763899	0.1585769	0.1215641	0.0868635
3 (2015-2019)	0.2158734	0.2033686	0.1759044	0.1590262	0.1213591	0.0868270

```

last_100 = tibble(RPS = apply(HDA[c(251:350),c(81:104)], 2, mean),
  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 4)),
  Model = factor(c(rep("0", 6), rep("3", 6), rep("3 (2019)", 6),
    rep("3 (2015-2019)", 6)),
    levels = c("0", "3", "3 (2019)", "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	0.2179326	0.1980936	0.1803926	0.1673810	0.1201216	0.0951005
3	0.2150895	0.1986115	0.1766849	0.1627764	0.1184260	0.0912985
3 (2019)	0.2117740	0.1957424	0.1730015	0.1601052	0.1168939	0.0902091
3 (2015-2019)	0.2109743	0.1943738	0.1735835	0.1613448	0.1169494	0.0911196