Brier Score

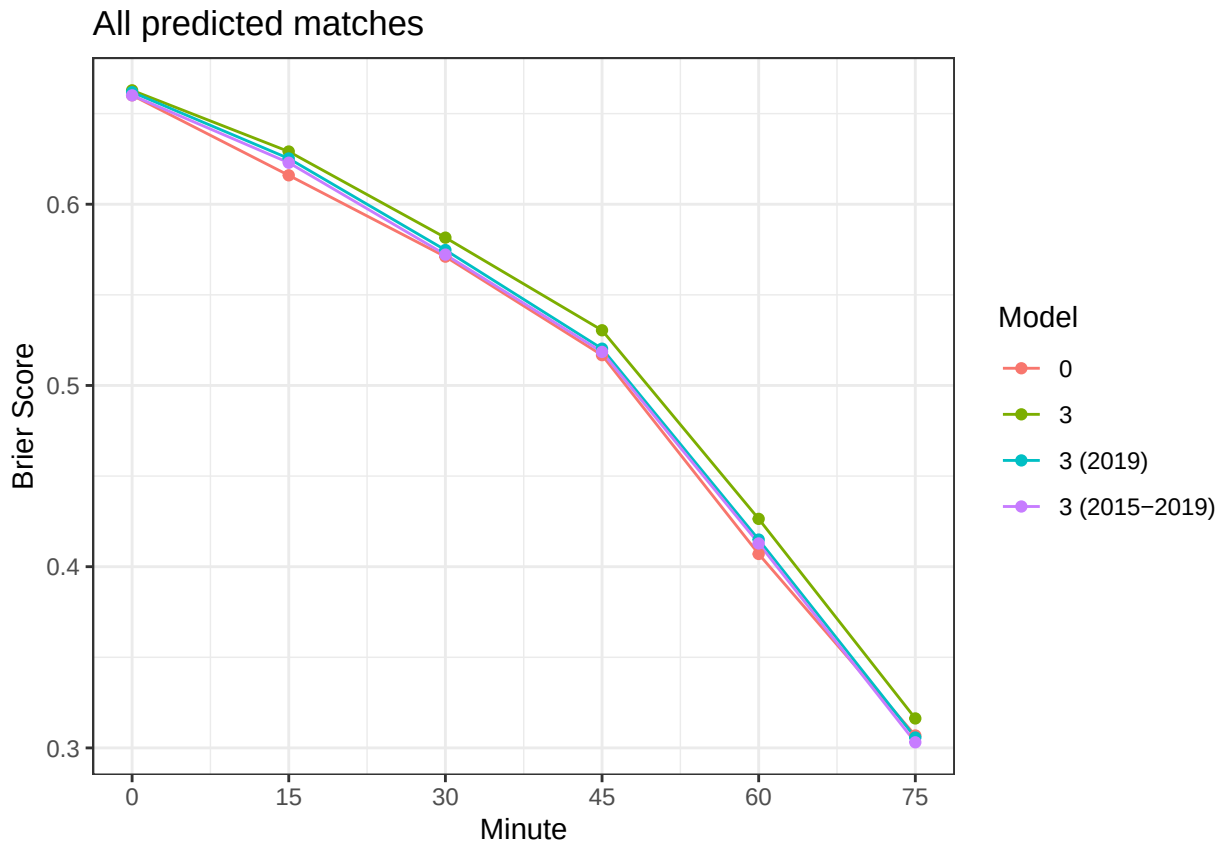
```
library(dplyr)
library(ggplot2)
library(tidyr)
library(knitr)

load("data/HDA.RData")

nrow(HDA)
```

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

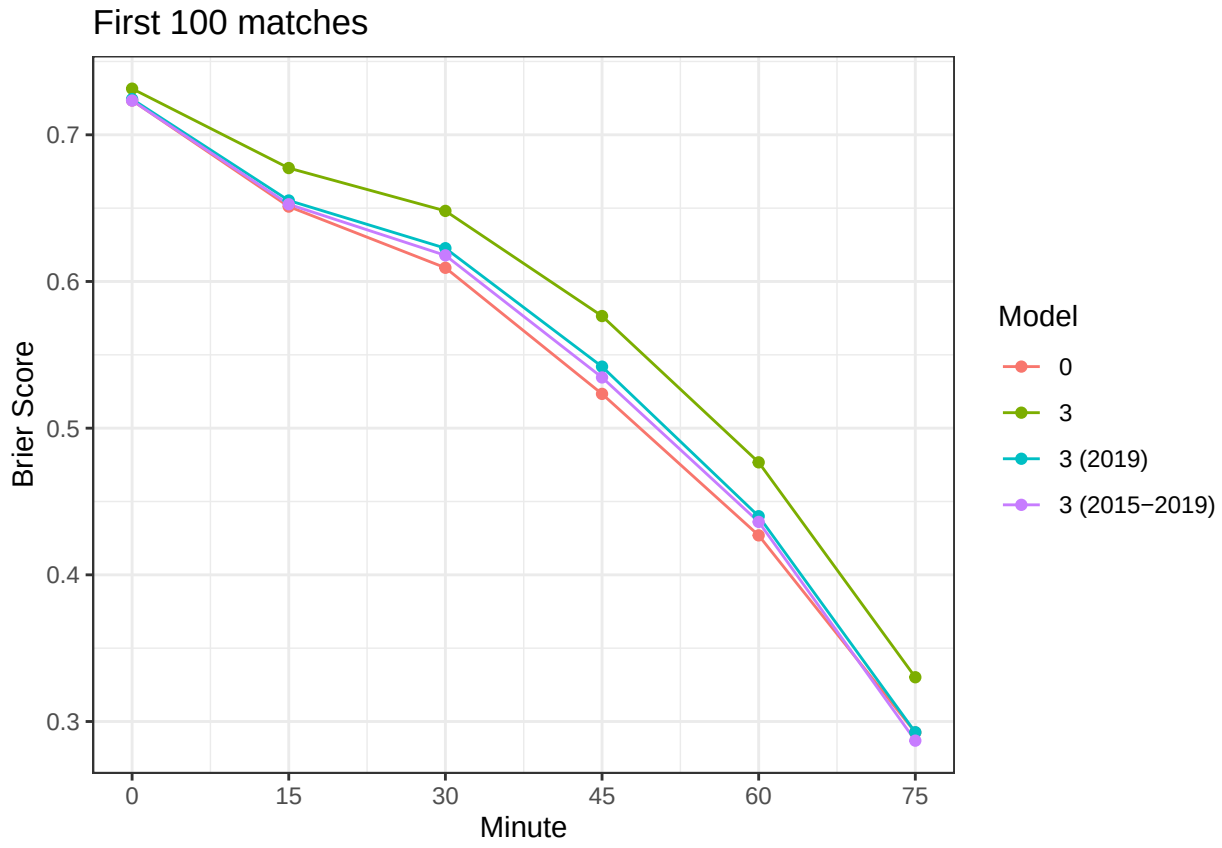
```
all = tibble(Brier = apply(HDA[,c(105:128)], 2, mean),
             Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 4)),
             Model = factor(c(rep("0", 6), rep("3", 6), rep("3 (2019)", 6),
                              rep("3 (2015-2019)", 6)),
                            levels = c("0", "3", "3 (2019)", "3 (2015-2019)")))
all %>%
  ggplot(aes(x = Minute, y = 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	0.6601204	0.6159520	0.5710751	0.5167246	0.4070241	0.3068499
3	0.6628437	0.6290749	0.5815726	0.5304530	0.4263846	0.3162705
3 (2019)	0.6618385	0.6252399	0.5747260	0.5202837	0.4149573	0.3057443
3 (2015-2019)	0.6600751	0.6229081	0.5721995	0.5185039	0.4128283	0.3031031

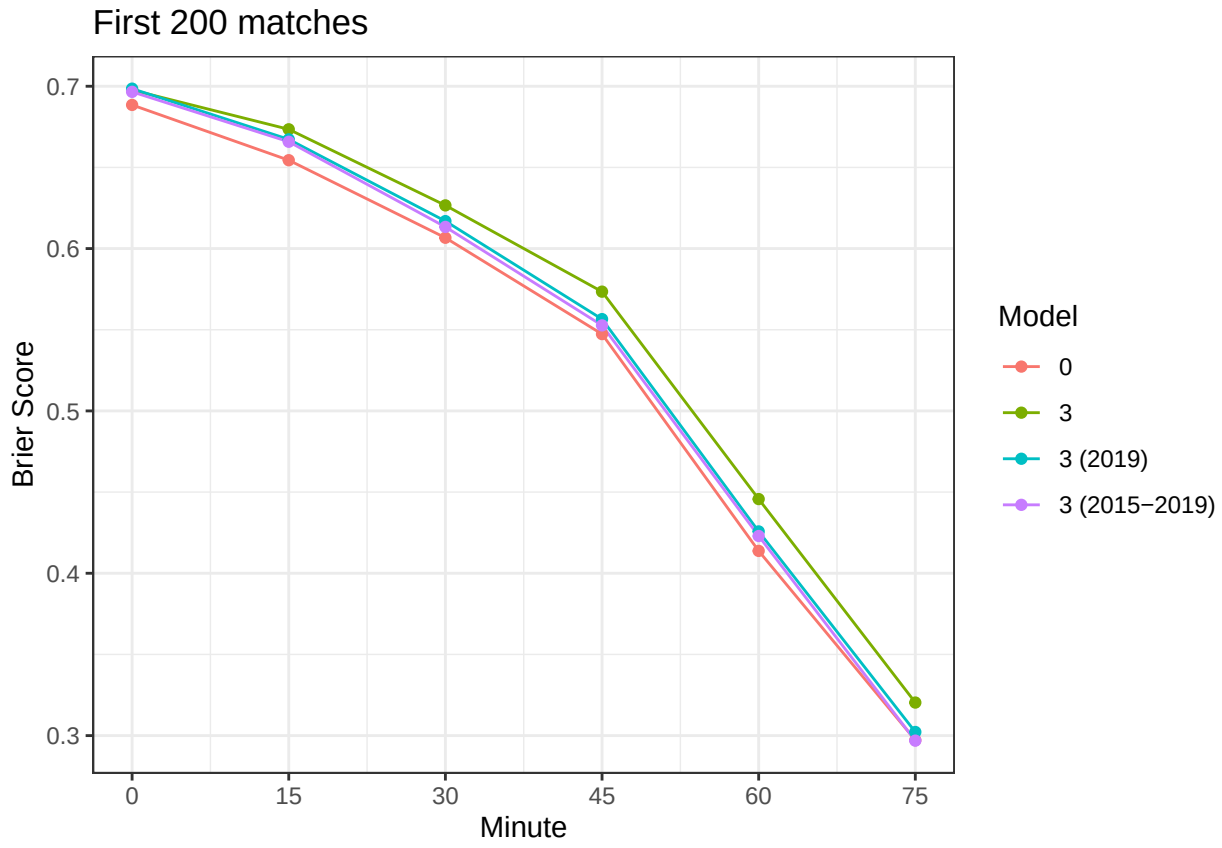
```
first_100 = tibble(Brier = apply(HDA[c(1:100),c(105:128)], 2, mean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 4)),
                   Model = factor(c(rep("0", 6), rep("3", 6), rep("3 (2019)", 6),
                                    rep("3 (2015-2019)", 6)),
                                  levels = c("0", "3", "3 (2019)", "3 (2015-2019)")))
first_100 %>%
  ggplot(aes(x = Minute, y = 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	0.7234222	0.6510753	0.6093454	0.5233602	0.4269015	0.2925227
3	0.7314636	0.6773334	0.6481162	0.5764371	0.4766991	0.3301632
3 (2019)	0.7243425	0.6552194	0.6226844	0.5419386	0.4399636	0.2926757
3 (2015-2019)	0.7233623	0.6526449	0.6178094	0.5345832	0.4361013	0.2869517

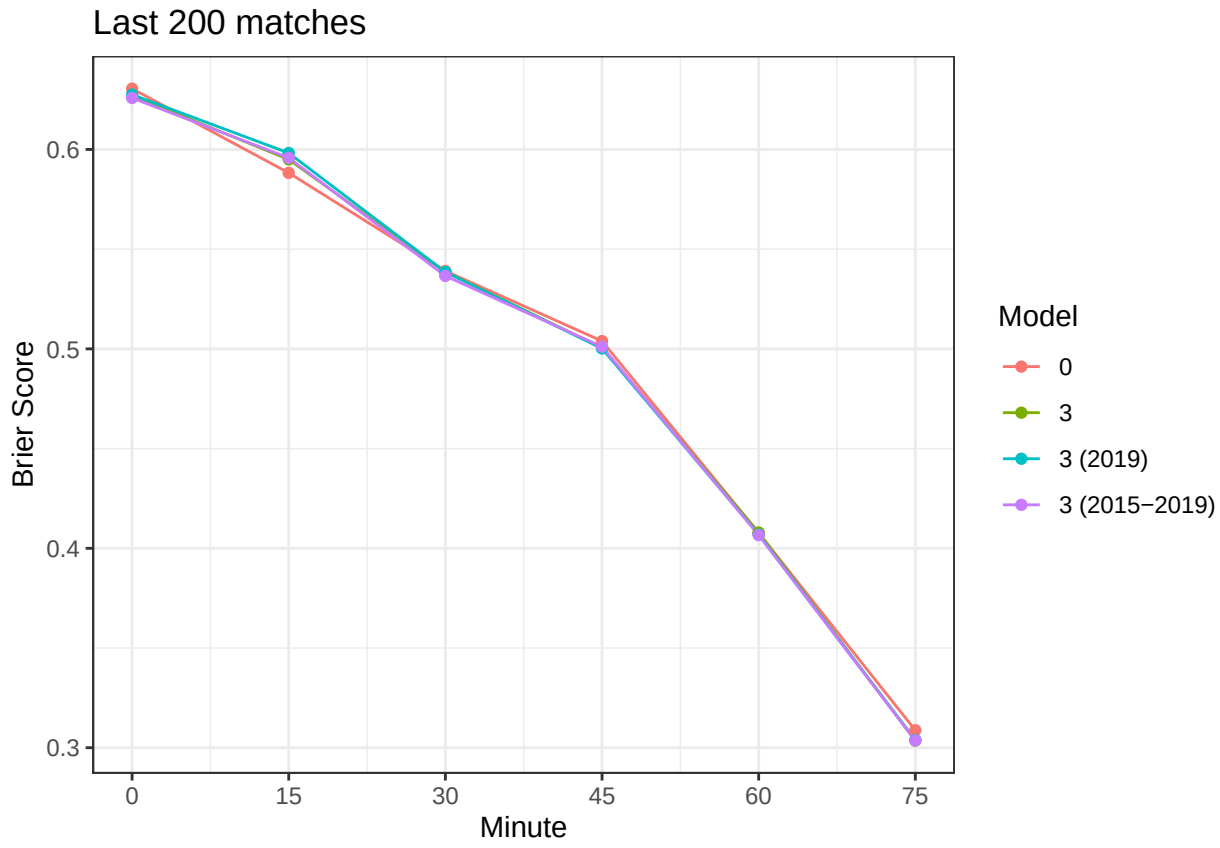
```
first_200 = tibble(Brier = apply(HDA[c(1:200),c(105:128)], 2, mean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 4)),
                   Model = factor(c(rep("0", 6), rep("3", 6), rep("3 (2019)", 6),
                                    rep("3 (2015-2019)", 6)),
                                  levels = c("0", "3", "3 (2019)", "3 (2015-2019)")))
first_200 %>%
  ggplot(aes(x = Minute, y = 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	0.6884814	0.6544402	0.6066652	0.5473290	0.4137692	0.2972172
3	0.6976919	0.6734674	0.6266114	0.5734665	0.4457118	0.3203625
3 (2019)	0.6985029	0.6672866	0.6169179	0.5565552	0.4257719	0.3022293
3 (2015-2019)	0.6965541	0.6658075	0.6133372	0.5526954	0.4229632	0.2969022

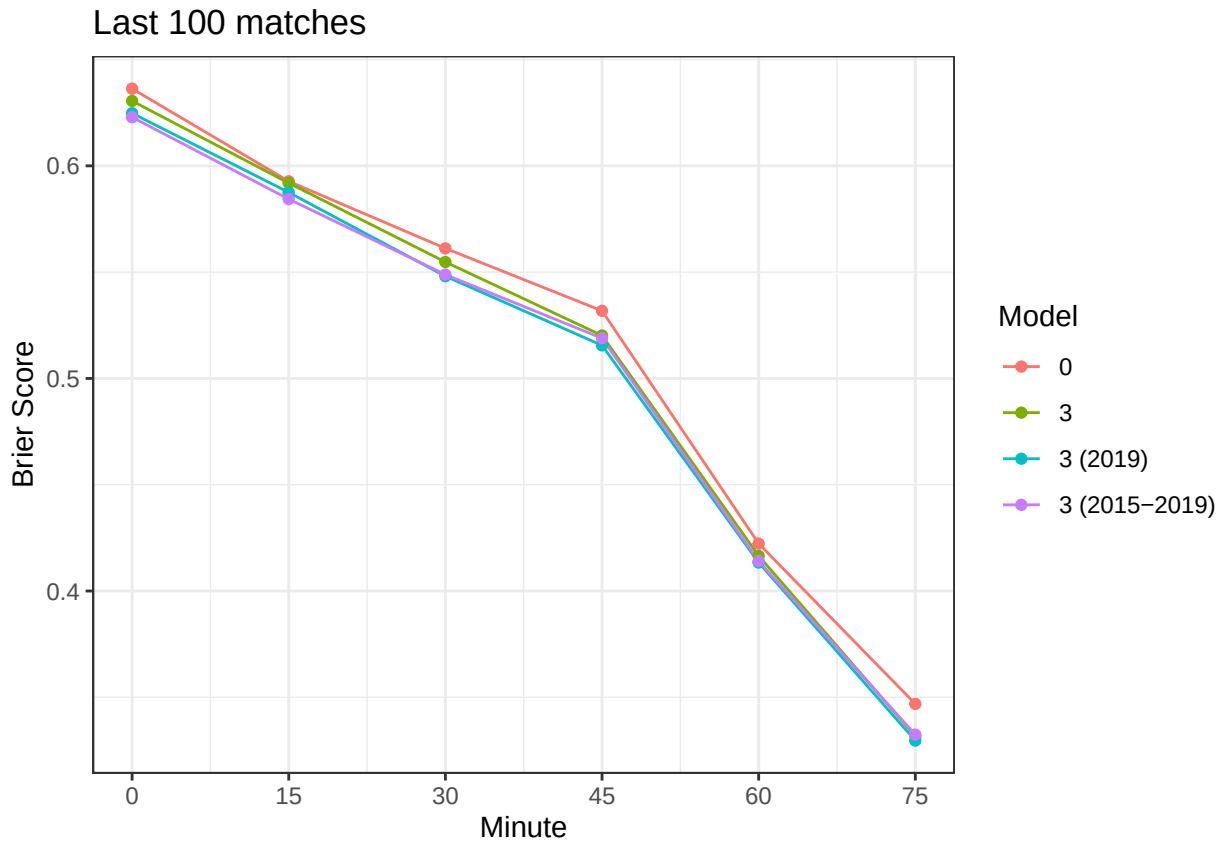
```
last_200 = tibble(Brier = apply(HDA[c(151:350),c(105:128)], 2, mean),
                  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 4)),
                  Model = factor(c(rep("0", 6), rep("3", 6), rep("3 (2019)", 6),
                                   rep("3 (2015-2019)", 6)),
                                 levels = c("0", "3", "3 (2019)", "3 (2015-2019)")))
last_200 %>%
  ggplot(aes(x = Minute, y = 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	0.6304837	0.5882607	0.5390675	0.5038946	0.4076944	0.3087446
3	0.6270157	0.5949916	0.5385244	0.5005590	0.4079133	0.3035479
3 (2019)	0.6275244	0.5982227	0.5385617	0.5001949	0.4069254	0.3039285
3 (2015-2019)	0.6258061	0.5957750	0.5365672	0.5010820	0.4066271	0.3036461

```
last_100 = tibble(Brier = apply(HDA[c(251:350),c(105:128)], 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 = 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	0.6363099	0.5927242	0.5611849	0.5317985	0.4223018	0.3468986
3	0.6305436	0.5919097	0.5547845	0.5201967	0.4165304	0.3316921
3 (2019)	0.6247647	0.5875229	0.5480837	0.5155581	0.4133862	0.3296110
3 (2015-2019)	0.6228720	0.5843654	0.5488402	0.5188887	0.4140975	0.3324337