Geometric mean for the scores

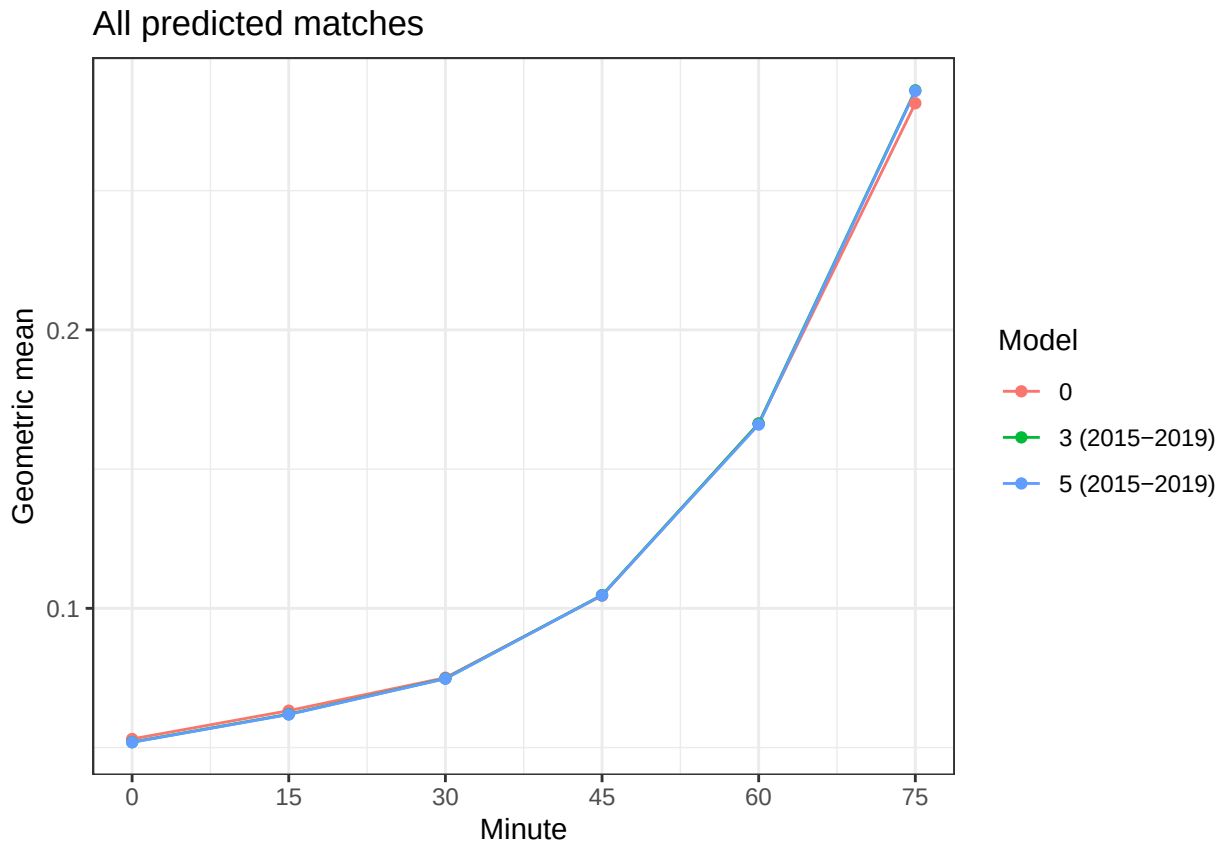
```
library(dplyr)
library(ggplot2)
library(tidyr)
library(knitr)

load("data/HDA2.RData")
load("~/GitHub/soccer-live-predictions/soccer-live-predictions/scrape/data/reds.RData")

nrow(HDA2)
```

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

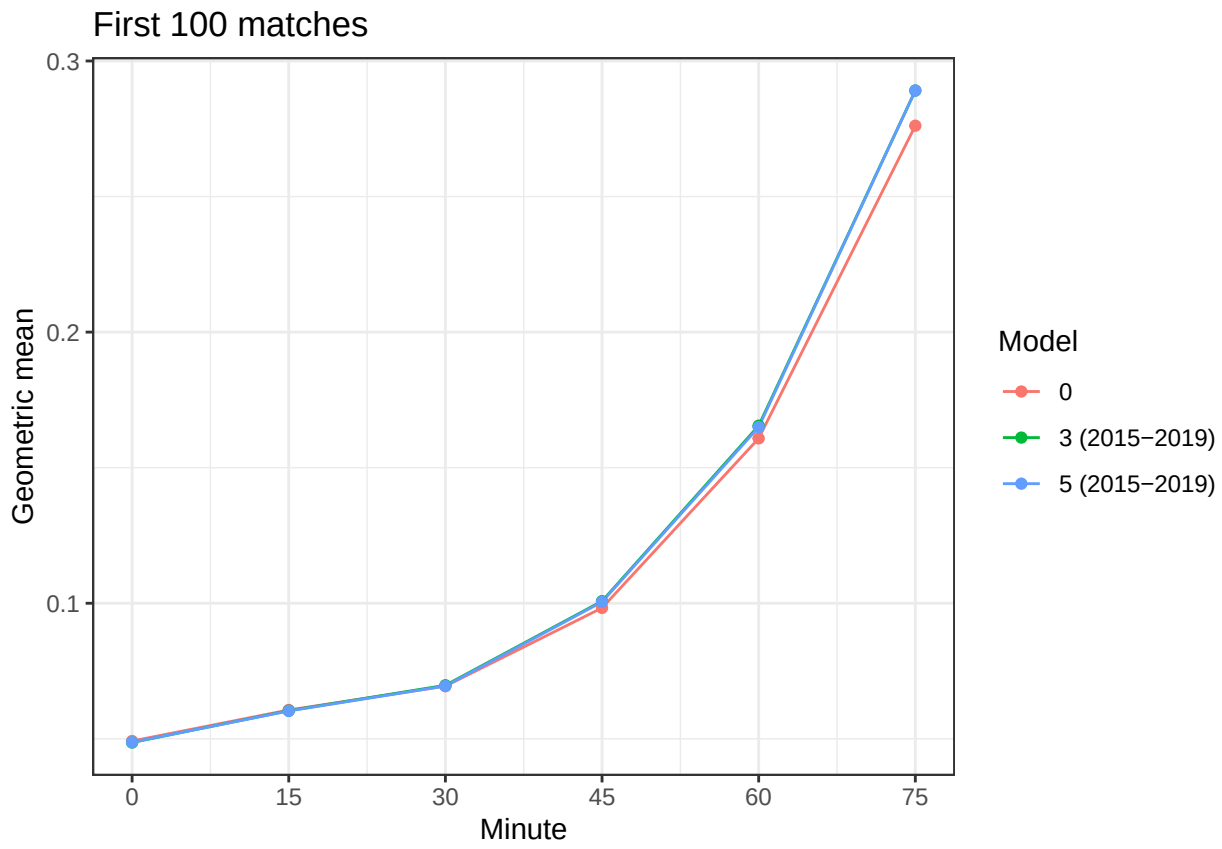
```
all = tibble(GeoMean = apply(HDA2[,c(219:224, 237:248)], 2,
                             EnvStats::geoMean),
             Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 3)),
             Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6),
                              rep("5 (2015-2019)", 6)),
                            levels = c("0", "3 (2015-2019)", "5 (2015-2019)")))
all %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches") +
  ylab("Geometric mean")
```




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

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.0530792	0.0632468	0.0750837	0.1045825	0.1663690	0.2813975
3 (2015-2019)	0.0519588	0.0619883	0.0748002	0.1047464	0.1663731	0.2860349
5 (2015-2019)	0.0518882	0.0618657	0.0747223	0.1046600	0.1660005	0.2857975

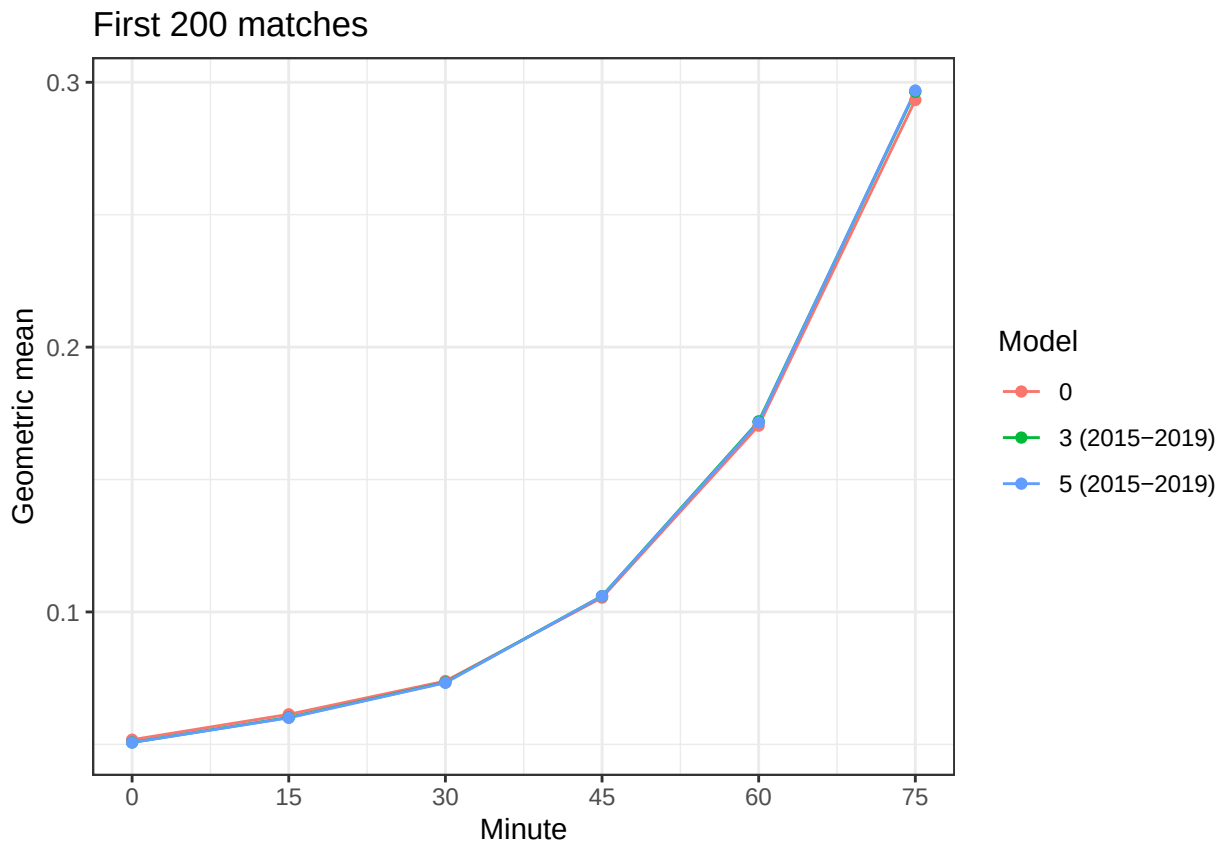
```
first_100 = tibble(GeoMean = apply(HDA2[c(1:100), c(219:224, 237:248)], 2,
                                   EnvStats::geoMean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 3)),
                   Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6),
                                    rep("5 (2015-2019)", 6)),
                                  levels = c("0", "3 (2015-2019)", "5 (2015-2019)")))
first_100 %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("First 100 matches") +
  ylab("Geometric mean")
```




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

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.0491718	0.0606188	0.0694235	0.0982771	0.1608002	0.2761246
3 (2015-2019)	0.0486287	0.0603747	0.0697429	0.1007715	0.1654847	0.2890992
5 (2015-2019)	0.0487278	0.0602696	0.0694816	0.1005074	0.1649647	0.2890969

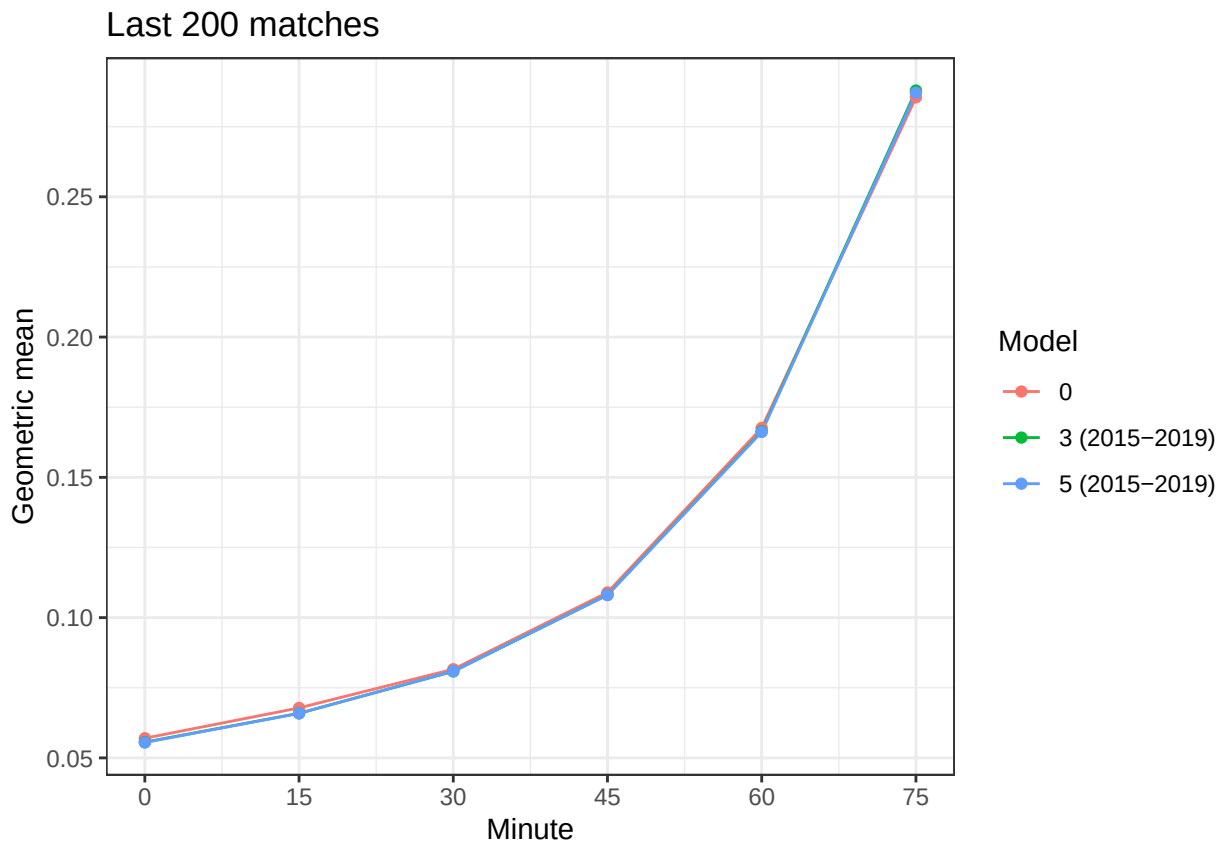
```
first_200 = tibble(GeoMean = apply(HDA2[c(1:200), c(219:224, 237:248)], 2,
                                   EnvStats::geoMean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 3)),
                   Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6),
                                    rep("5 (2015-2019)", 6)),
                                  levels = c("0", "3 (2015-2019)", "5 (2015-2019)")))
first_200 %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("First 200 matches") +
  ylab("Geometric mean")
```




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

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.0517575	0.0613402	0.0738980	0.1054137	0.1703355	0.2932999
3 (2015-2019)	0.0507330	0.0601672	0.0734091	0.1059920	0.1719712	0.2964352
5 (2015-2019)	0.0506983	0.0600202	0.0732535	0.1059688	0.1715673	0.2968768

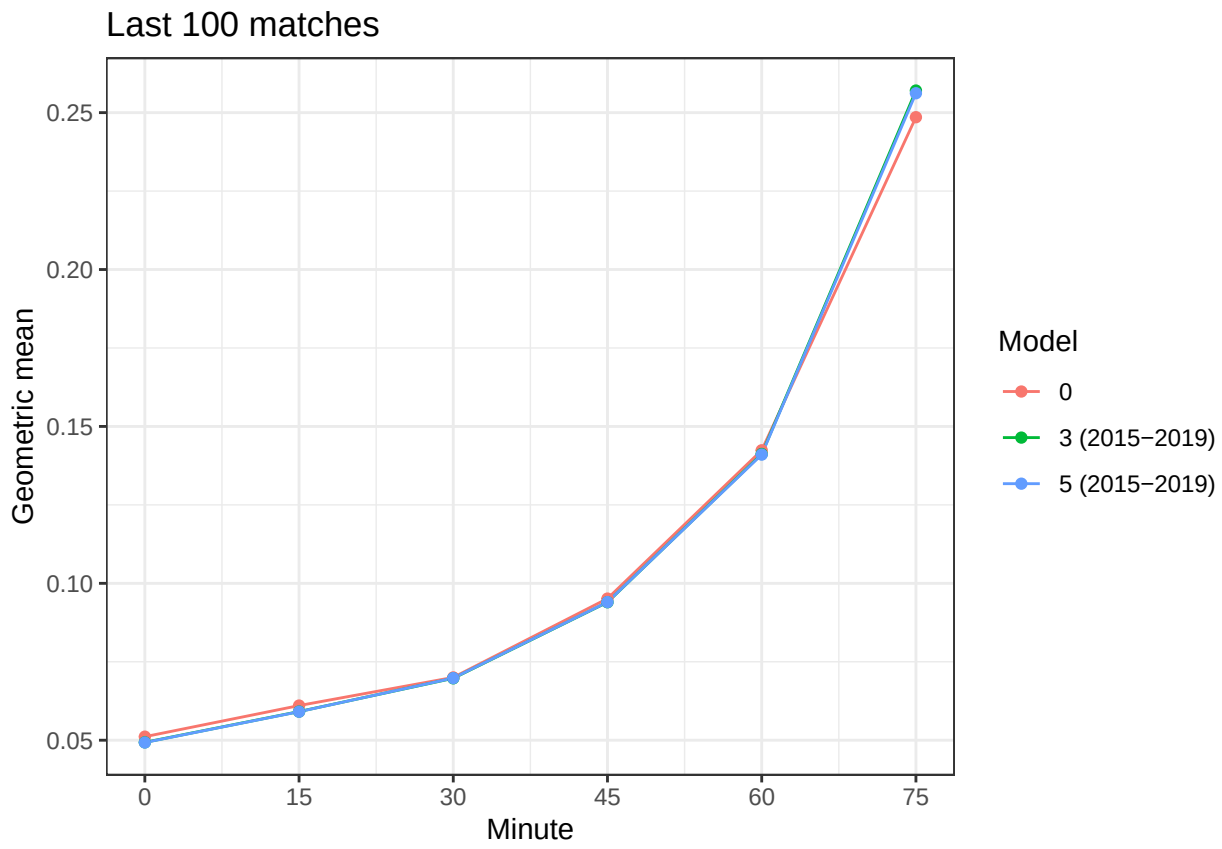
```
last_200 = tibble(GeoMean = apply(HDA2[c(134:333), c(219:224, 237:248)], 2,
                                  EnvStats::geoMean),
                  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 3)),
                  Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6),
                                   rep("5 (2015-2019)", 6)),
                                 levels = c("0", "3 (2015-2019)", "5 (2015-2019)")))
last_200 %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("Last 200 matches") +
  ylab("Geometric mean")
```




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

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.0570081	0.0678132	0.0816000	0.1089679	0.1676214	0.2854587
3 (2015-2019)	0.0556095	0.0658973	0.0808215	0.1080906	0.1664050	0.2878740
5 (2015-2019)	0.0554804	0.0658625	0.0808981	0.1080656	0.1661799	0.2870818

```
last_100 = tibble(GeoMean = apply(HDA2[c(234:333), c(219:224, 237:248)], 2,
                                  EnvStats::geoMean),
                  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 = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("Last 100 matches") +
  ylab("Geometric mean")
```




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

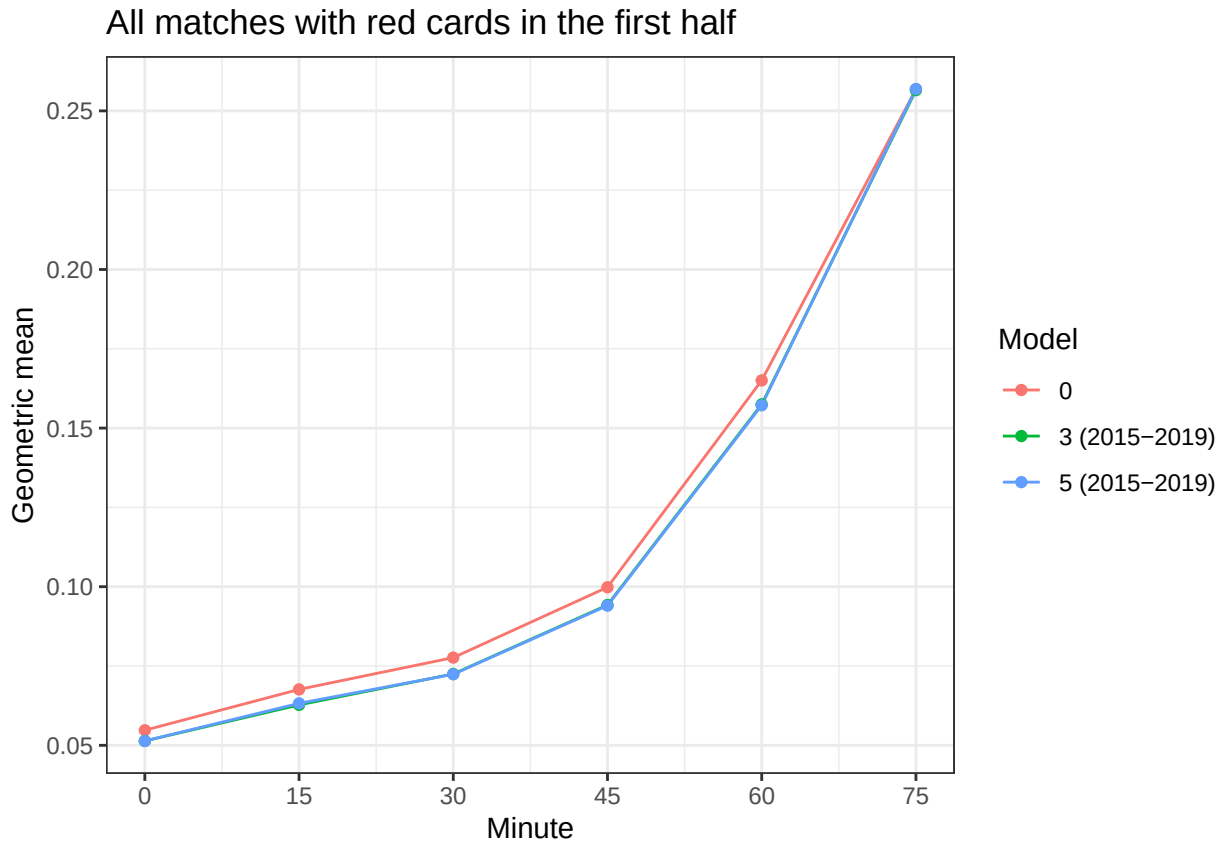
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.0511255	0.0610438	0.0700116	0.0951357	0.1423988	0.2485370
3 (2015-2019)	0.0493260	0.0591259	0.0697263	0.0939653	0.1412134	0.2570346
5 (2015-2019)	0.0492675	0.0590832	0.0697911	0.0940185	0.1410176	0.2561776

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

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

```
HDA2_reds = HDA2 %>%
  filter(Match %in% matches)

all_reds = tibble(GeoMean = apply(HDA2_reds[,c(219:224, 237:248)], 2,
                                  EnvStats::geoMean),
                  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 = GeoMean, 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("Geometric mean")
```



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

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.0547517	0.0676315	0.0776510	0.0998322	0.1650381	0.2568238
3 (2015-2019)	0.0513843	0.0627481	0.0724910	0.0942352	0.1574875	0.2565276
5 (2015-2019)	0.0513693	0.0632306	0.0724079	0.0940419	0.1571803	0.2568855