Geometric mean for the scores

```r
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)
```
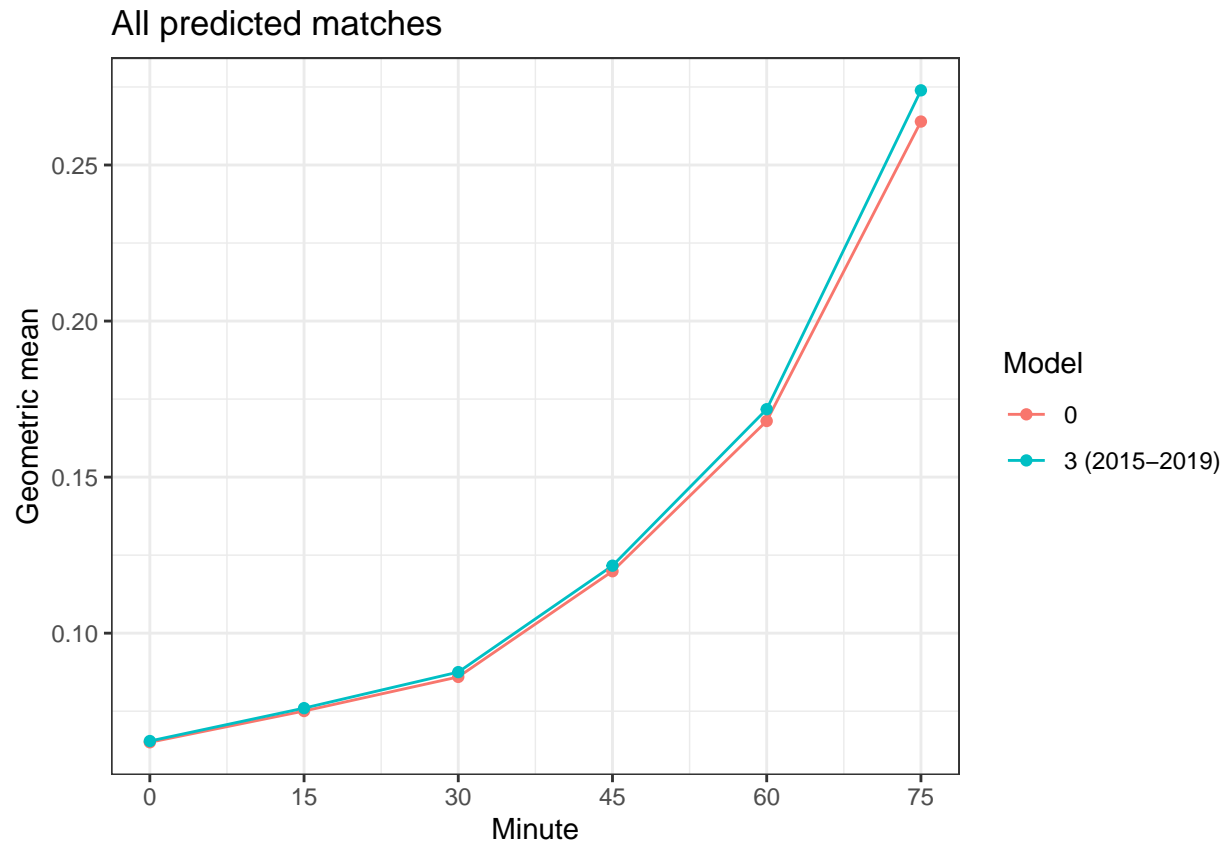
```
## [1] 340
```

```r
all = tibble(GeoMean = apply(HDA2[,c(93:104)], 2, EnvStats::geoMean),
             Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
             Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                            levels = c("0", "3 (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 predicted matches


```
all %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```
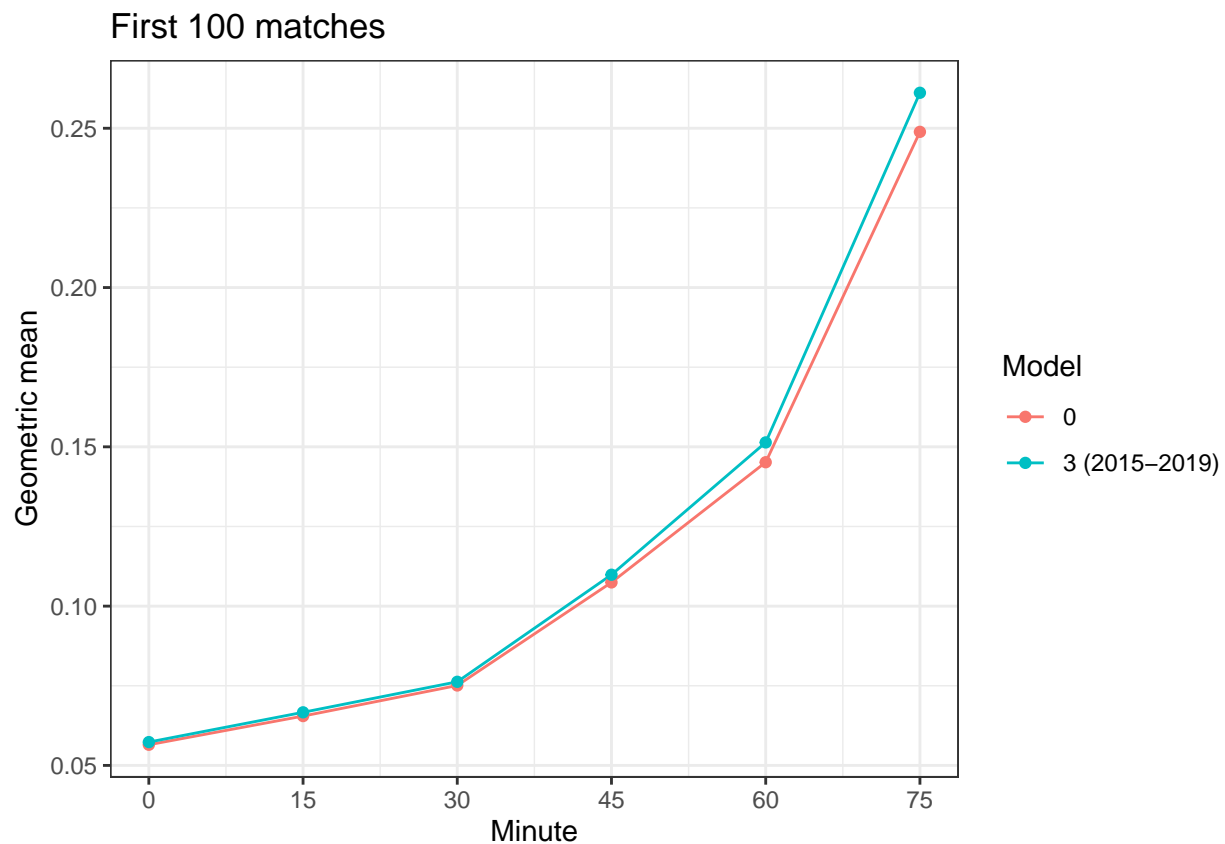
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.0650316	0.0750560	0.0859776	0.1198440	0.1679815	0.2639235
3 (2015-2019)	0.0654376	0.0759805	0.0875201	0.1216192	0.1717556	0.2739251

```
first_100 = tibble(GeoMean = apply(HDA2[c(1:100), c(93:104)], 2, EnvStats::geoMean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                   Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                                  levels = c("0", "3 (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()
```
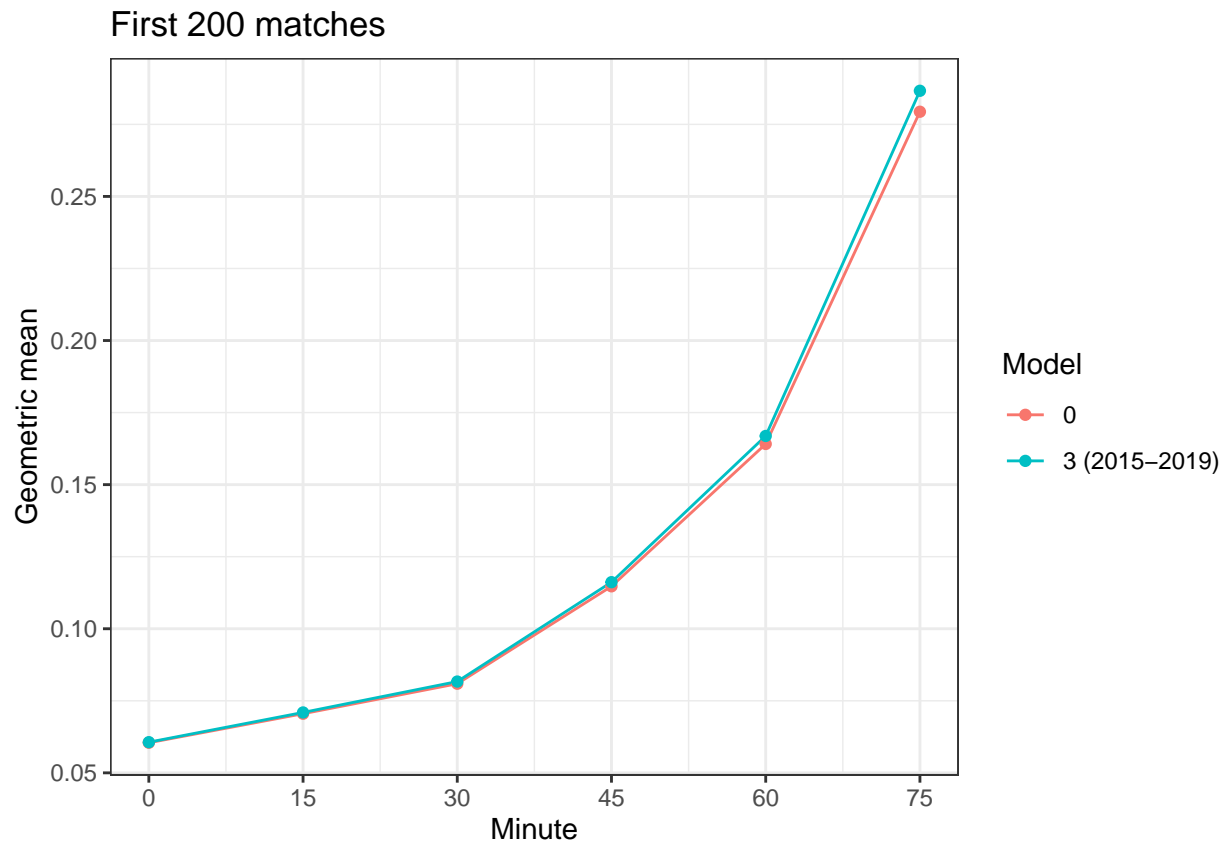
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.0564832	0.0654885	0.0750900	0.1074435	0.1451658	0.2488712
3 (2015-2019)	0.0573244	0.0666658	0.0762462	0.1098529	0.1513897	0.2611334

```
first_200 = tibble(GeoMean = apply(HDA2[c(1:200), c(93:104)], 2, EnvStats::geoMean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                   Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                               levels = c("0", "3 (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")
```
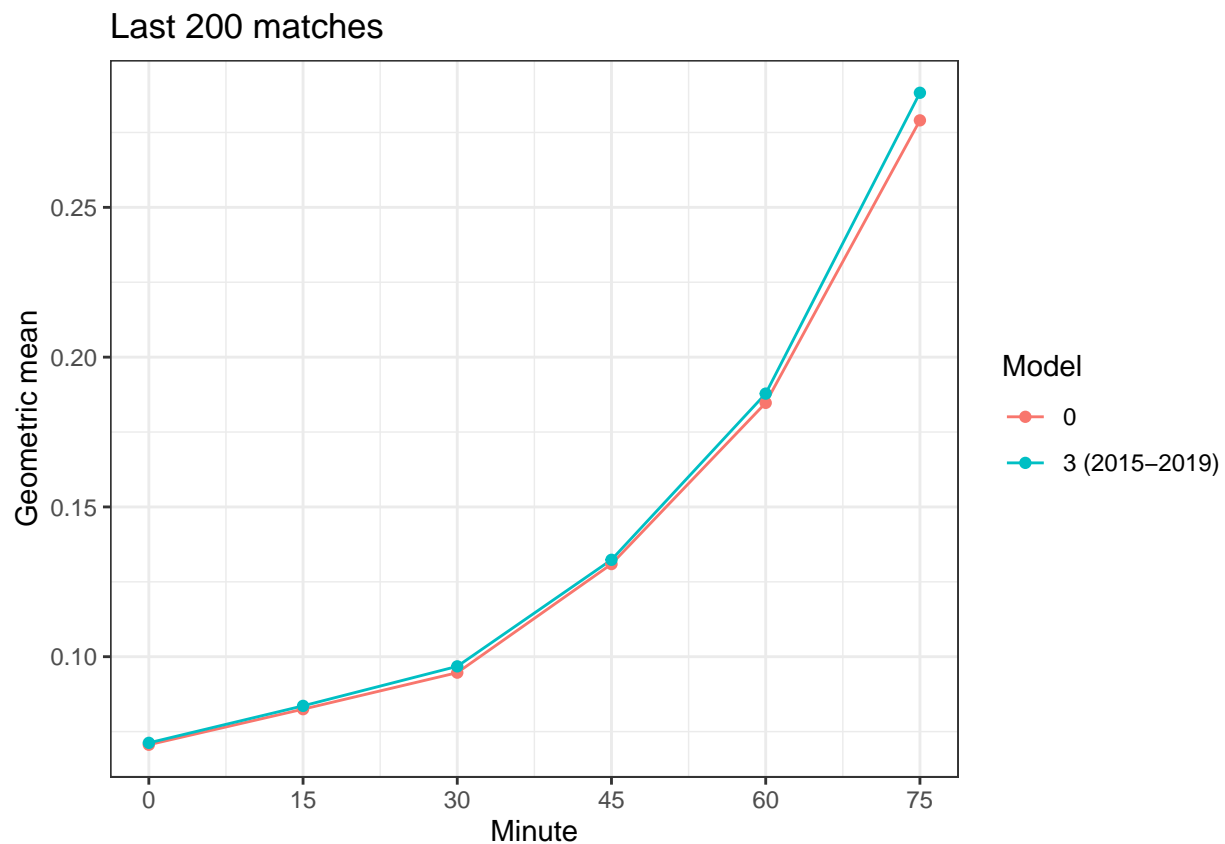
First 200 matches


```
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	0.0604376	0.0705065	0.0809200	0.1147200	0.1641100	0.2793880
3 (2015-2019)	0.0606382	0.0709494	0.0816766	0.1161591	0.1668909	0.2866547

```
last_200 = tibble(GeoMean = apply(HDA2[c(141:340), c(93:104)], 2, EnvStats::geoMean),
                  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                  Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                                 levels = c("0", "3 (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")
```
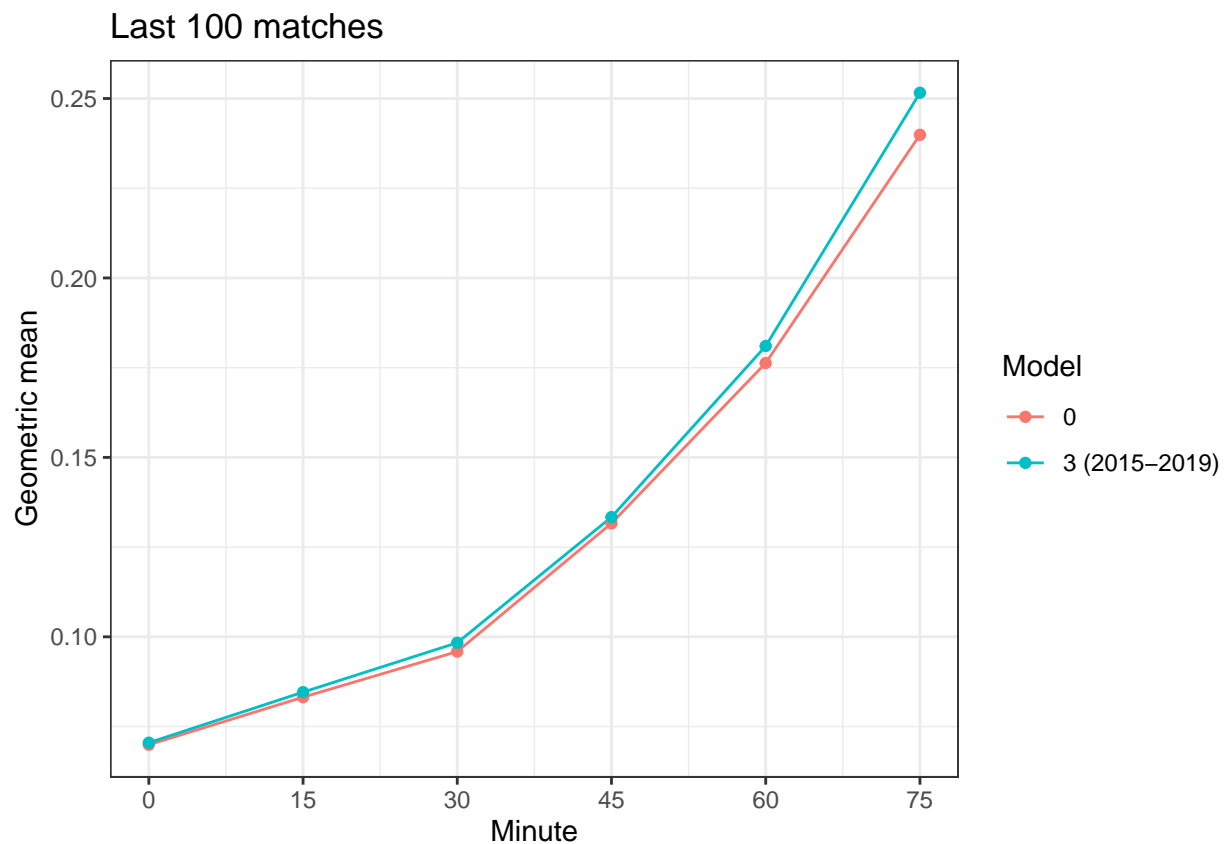

Last 200 matches

```
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	0.0706047	0.0824812	0.0946724	0.1309227	0.1847329	0.2790476
3 (2015-2019)	0.0712437	0.0835846	0.0967781	0.1323487	0.1878283	0.2882457

```
last_100 = tibble(GeoMean = apply(HDA2[c(241:340), c(93:104)], 2, EnvStats::geoMean),
                  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                  Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                                 levels = c("0", "3 (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()
```
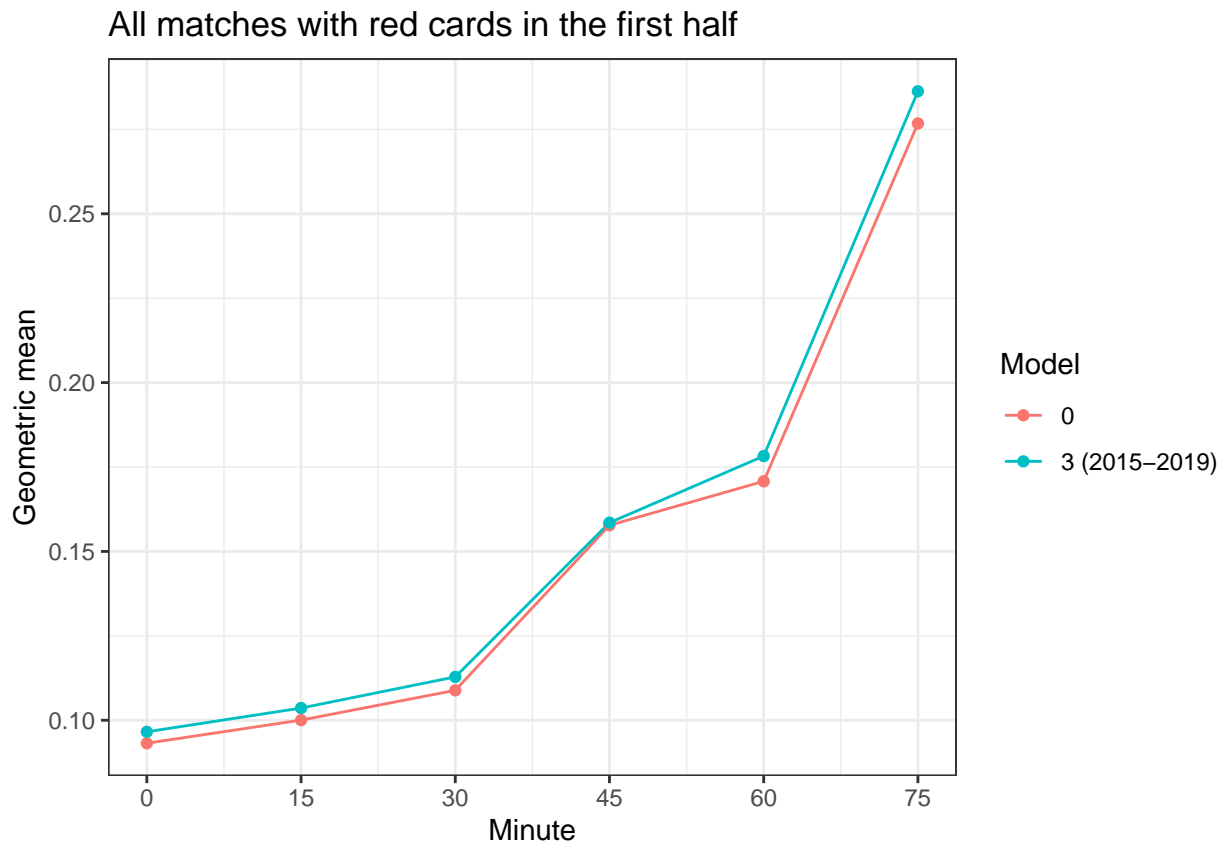
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.0699251	0.0831808	0.0958921	0.1316240	0.1762862	0.2398858
3 (2015-2019)	0.0704991	0.0845869	0.0983478	0.1333618	0.1810411	0.2515932

```
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(93:104)], 2, EnvStats::geoMean),
                  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                  Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                                 levels = c("0", "3 (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 matches with red cards in the first half

```
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	0.0931993	0.1000391	0.1088633	0.1577531	0.1707697	0.2767597
3 (2015-2019)	0.0965814	0.1036160	0.1128434	0.1585143	0.1782452	0.2862877