Geometric mean for the results

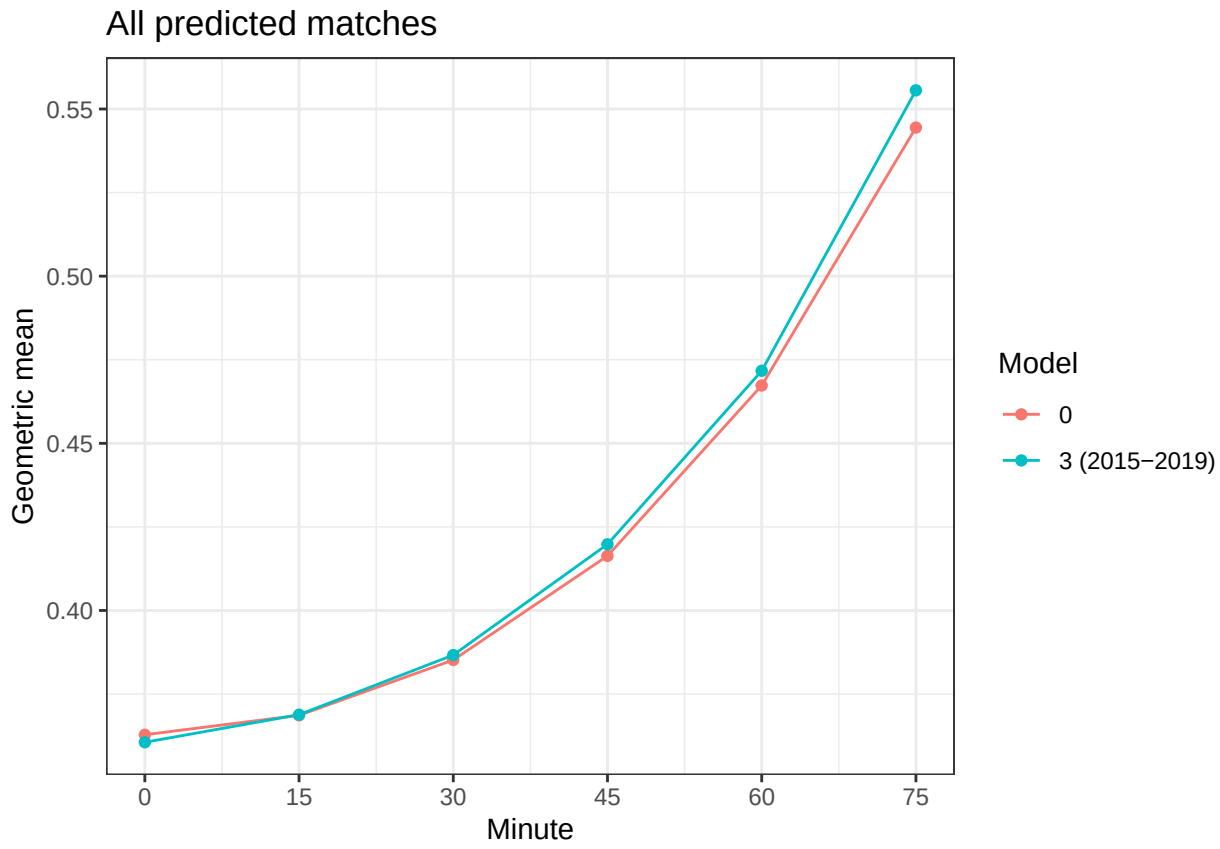
```
library(dplyr)
library(ggplot2)
library(tidyr)
library(knitr)

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

nrow(HDA)
```

```
## [1] 340
```

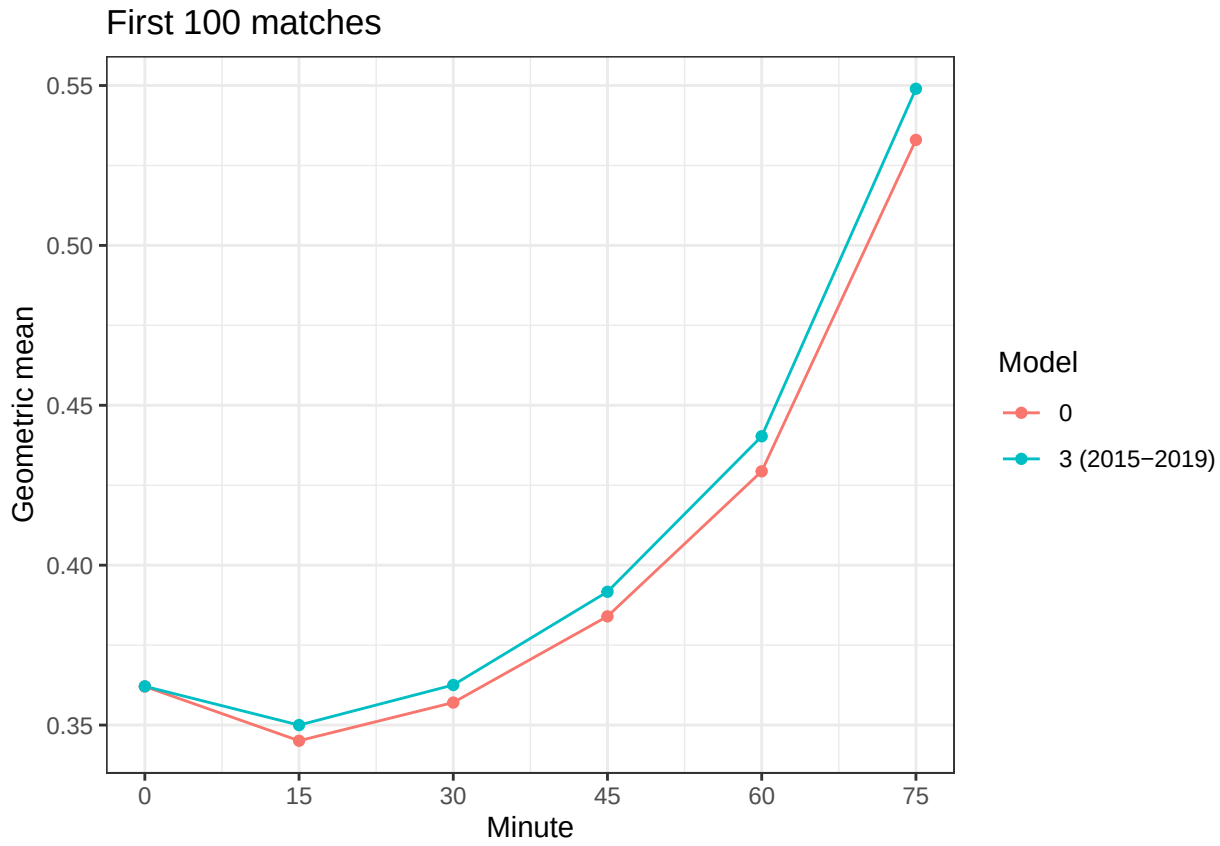
```
all = tibble(RPS = apply(HDA[,c(69:80)], 2, EnvStats::geoMean),
             Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
             Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                            levels = c("0", "3 (2015-2019)")))
all %>%
  ggplot(aes(x = Minute, y = 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("Geometric mean")
```




```
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	0.3628287	0.3686441	0.3851908	0.4163251	0.4672914	0.5444580
3 (2015-2019)	0.3605717	0.3688301	0.3866697	0.4198023	0.4716948	0.5555838

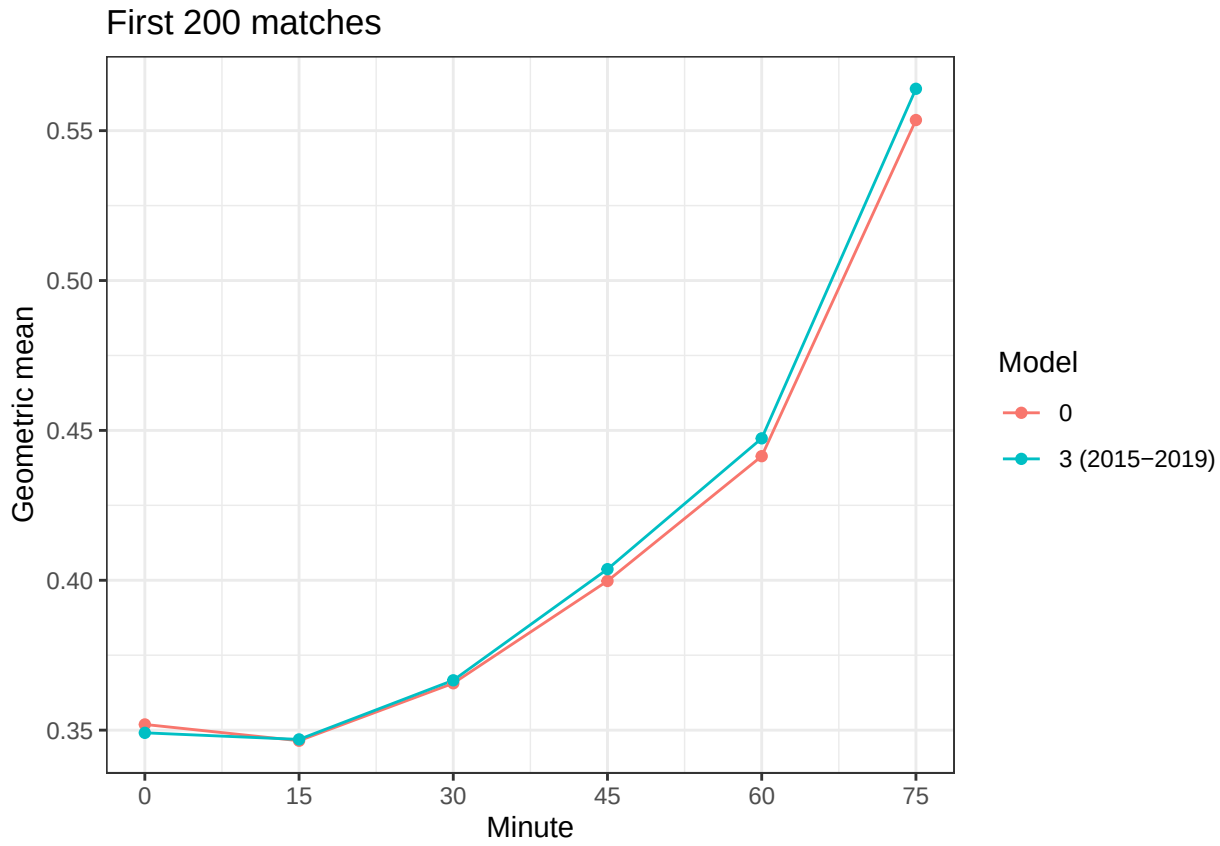
```
first_100 = tibble(RPS = apply(HDA[c(1:100), c(69:80)], 2, EnvStats::geoMean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                   Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                                  levels = c("0", "3 (2015-2019)")))
first_100 %>%
  ggplot(aes(x = Minute, y = 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("Geometric mean")
```




```
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	0.3620538	0.3450911	0.3570440	0.3839937	0.4293583	0.5329907
3 (2015-2019)	0.3621181	0.3499922	0.3625312	0.3916832	0.4403141	0.5489749

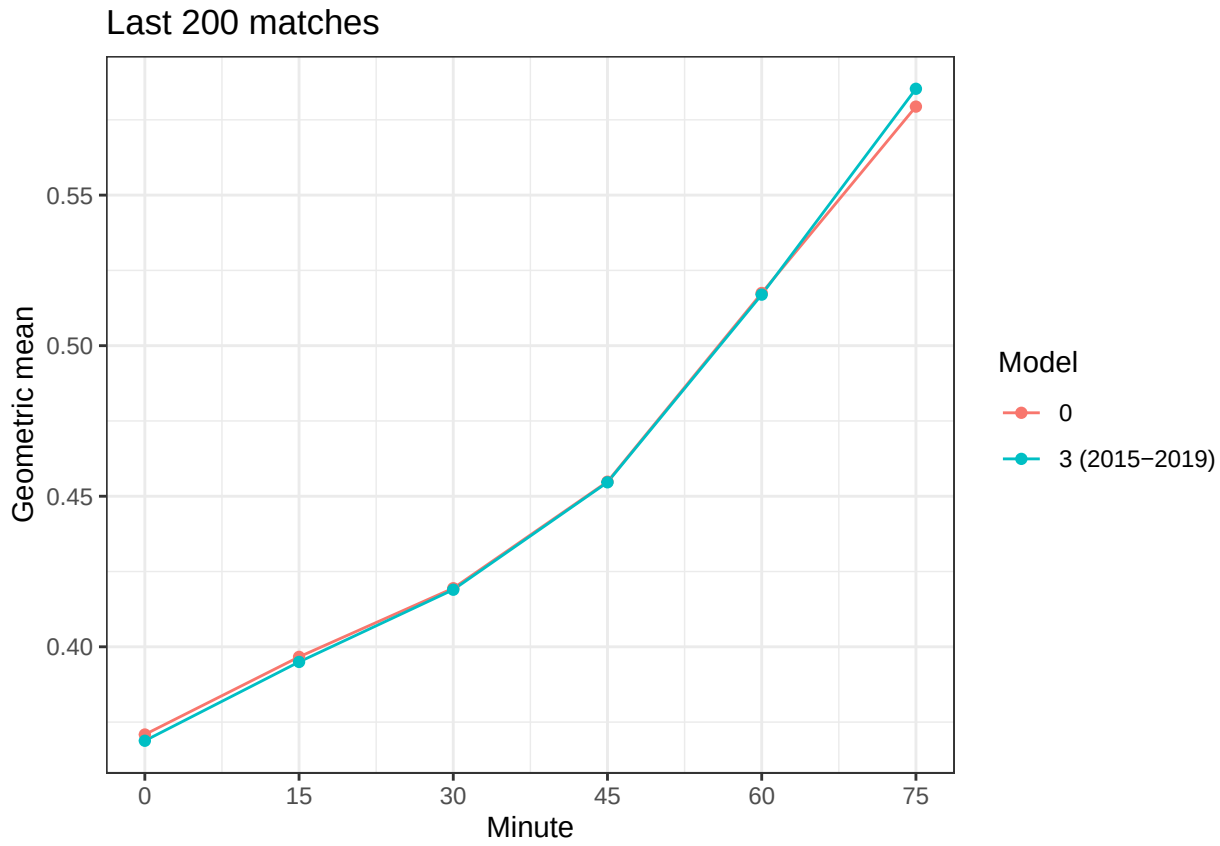
```
first_200 = tibble(RPS = apply(HDA[c(1:200), c(69:80)], 2, EnvStats::geoMean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                   Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                                  levels = c("0", "3 (2015-2019)")))
first_200 %>%
  ggplot(aes(x = Minute, y = 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("Geometric mean")
```




```
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	0.3519021	0.3464583	0.3656268	0.3997649	0.4414023	0.5535472
3 (2015-2019)	0.3491037	0.3469082	0.3666302	0.4037038	0.4473447	0.5639694

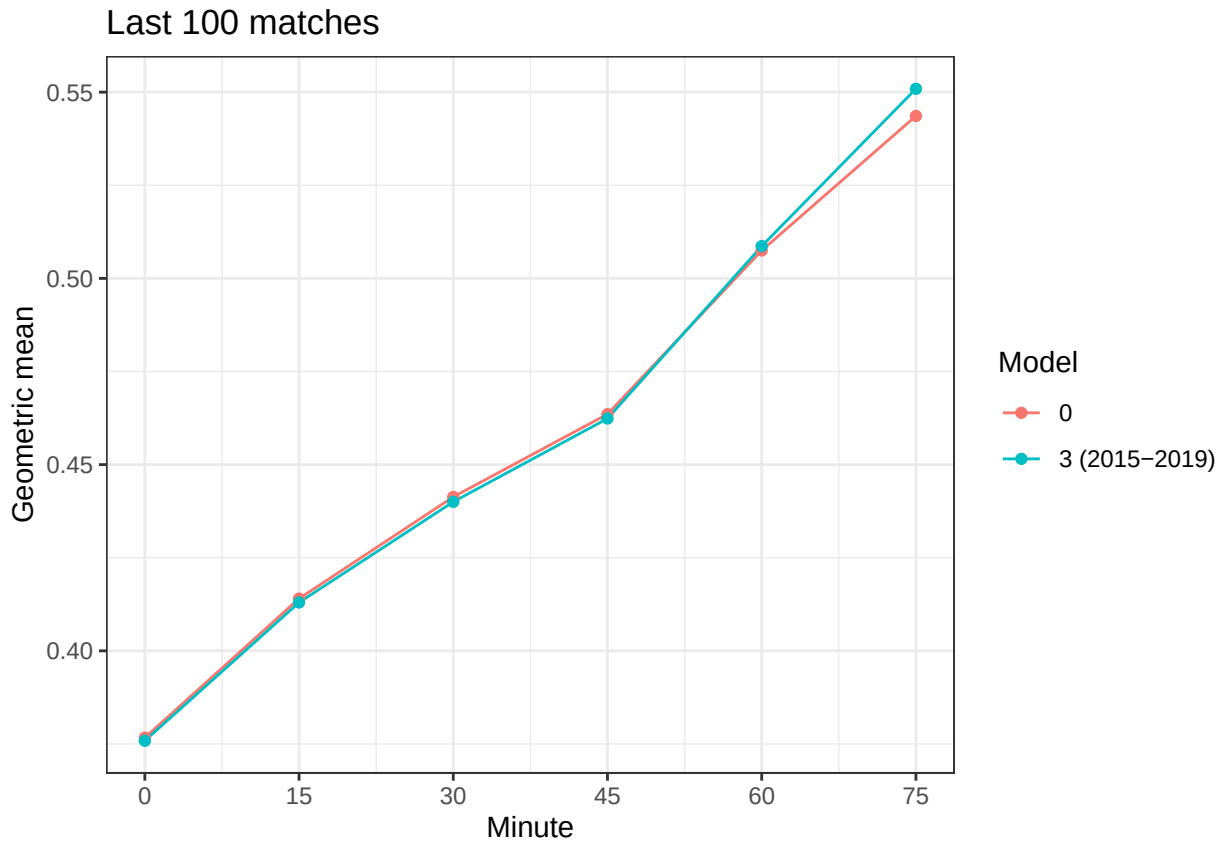
```
last_200 = tibble(RPS = apply(HDA[c(141:340), c(69:80)], 2, EnvStats::geoMean),
                  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                  Model = factor(c(rep("0", 6), rep("3 (2015-2019)", 6)),
                                 levels = c("0", "3 (2015-2019)")))
last_200 %>%
  ggplot(aes(x = Minute, y = 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("Geometric mean")
```




```
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	0.3708699	0.3966564	0.4194588	0.4548366	0.5174924	0.5794002
3 (2015-2019)	0.3688013	0.3949765	0.4189616	0.4546306	0.5169551	0.5853095

```
last_100 = tibble(RPS = apply(HDA[c(241:340), c(69:80)], 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 = 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("Geometric mean")
```




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

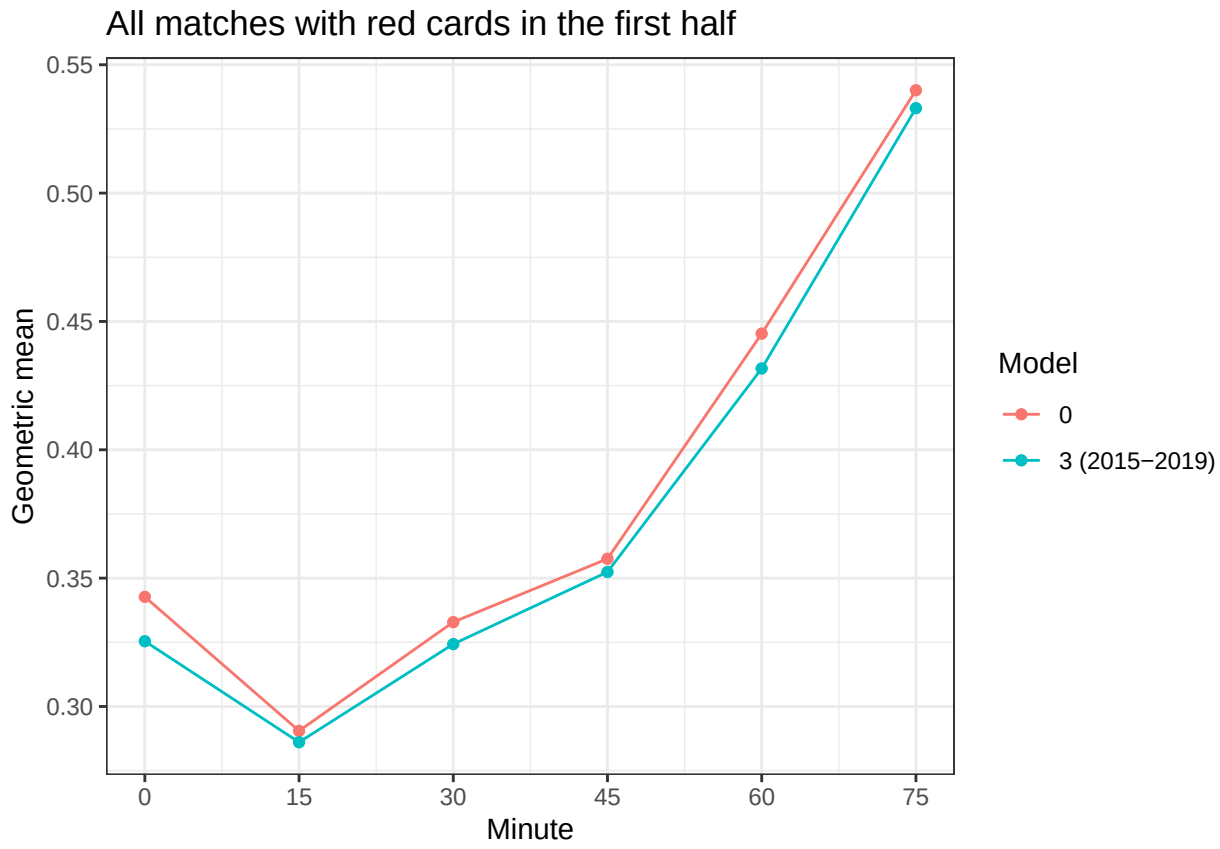
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.3766885	0.4140283	0.4413200	0.4635514	0.5074994	0.5436051
3 (2015-2019)	0.3758583	0.4129862	0.4400117	0.4623742	0.5086552	0.5509048

```
matches = reds %>%
  filter(Season == 2019, Half == 1) %>%
  .$Match
length(matches)
```

```
## [1] 17
```

```
HDA_reds = HDA %>%
  filter(Match %in% matches)

all_reds = tibble(RPS = apply(HDA_reds[,c(69:80)], 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 = 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("Geometric mean")
```



```
all_reds %>%
  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	0.3427094	0.2904955	0.3328753	0.3575332	0.4452553	0.5400644
3 (2015-2019)	0.3254306	0.2860743	0.3243052	0.3524055	0.4316814	0.5330754