Geometric mean for the results

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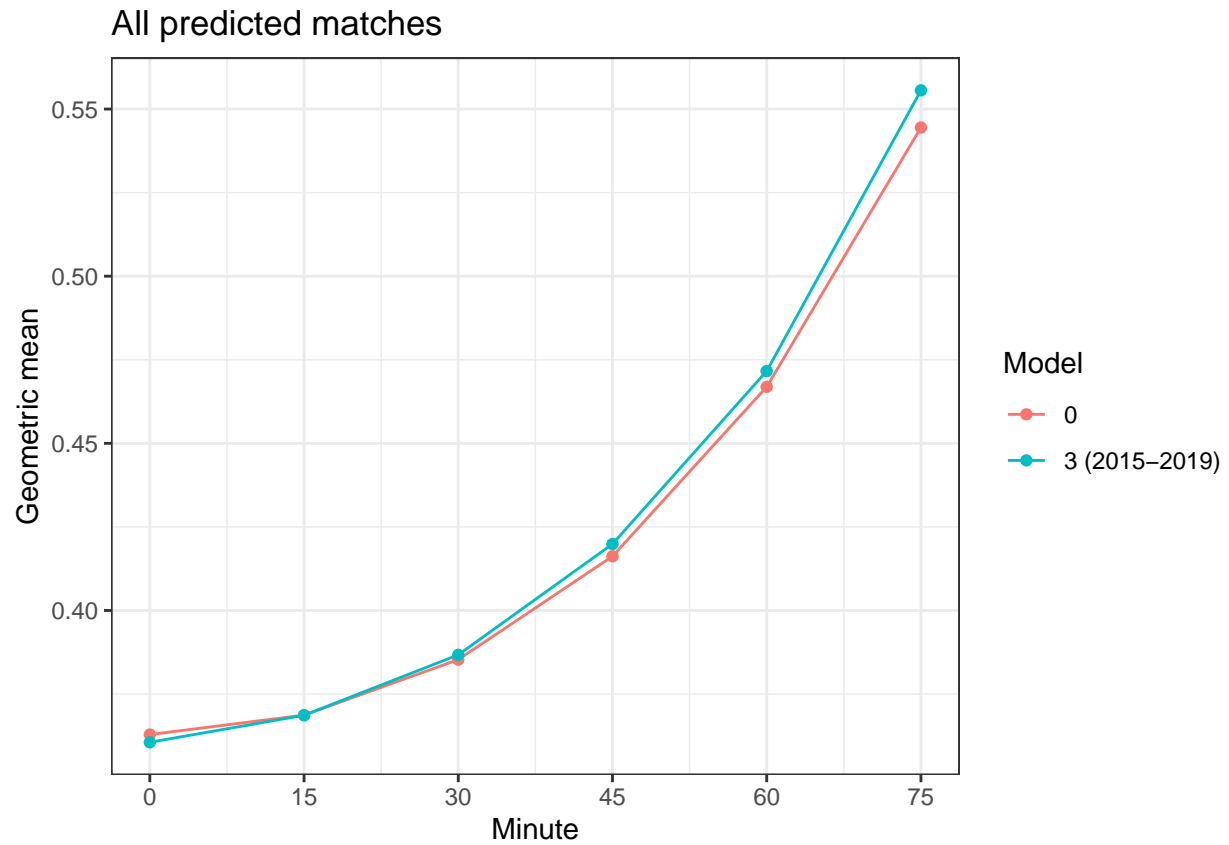
```
## [1] 340
```

```r
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")
```
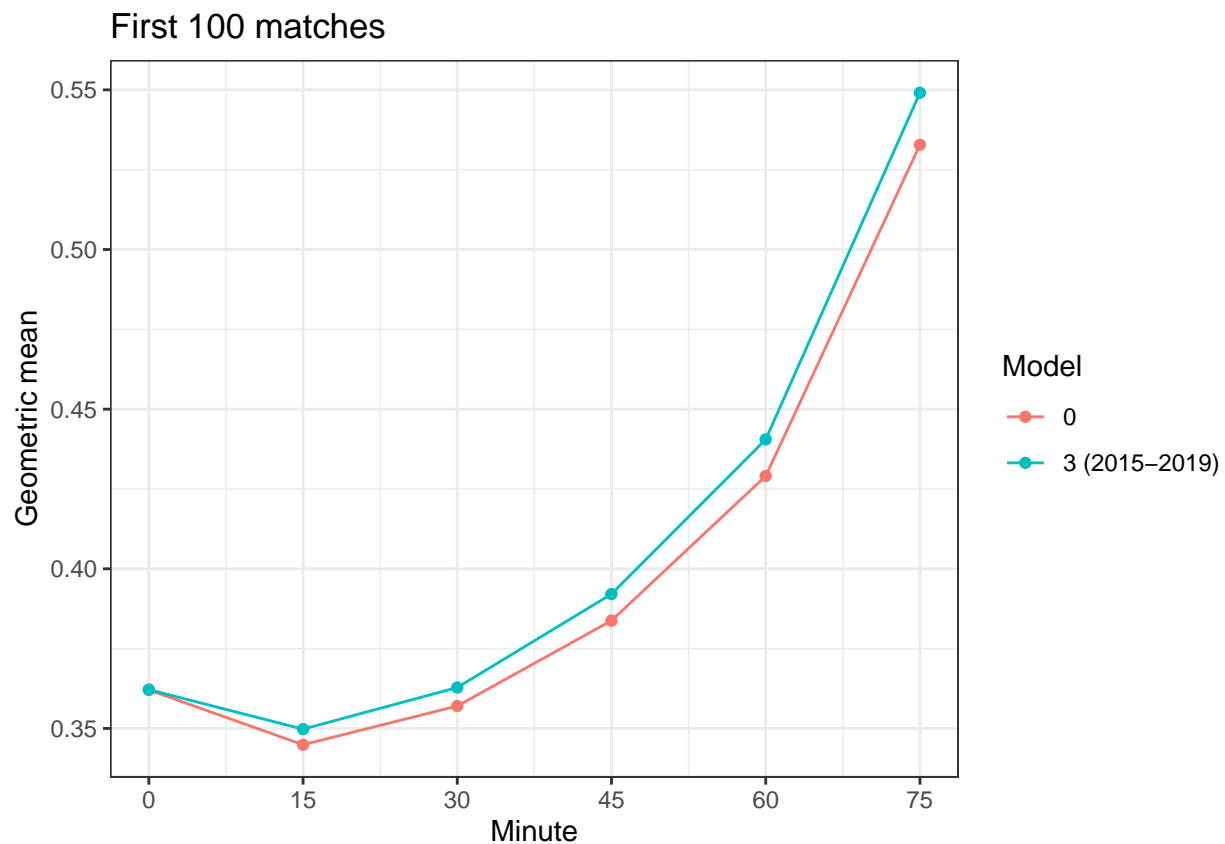
All predicted matches


```
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	0.3628867	0.3686914	0.3853030	0.4162097	0.4669075	0.5444719
3 (2015-2019)	0.3605656	0.3686503	0.3867329	0.4198765	0.4716247	0.5555977

```
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 matches

```
first_100 %>%
  pivot_wider(id_cols = "Model", values_from = "RPS", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```
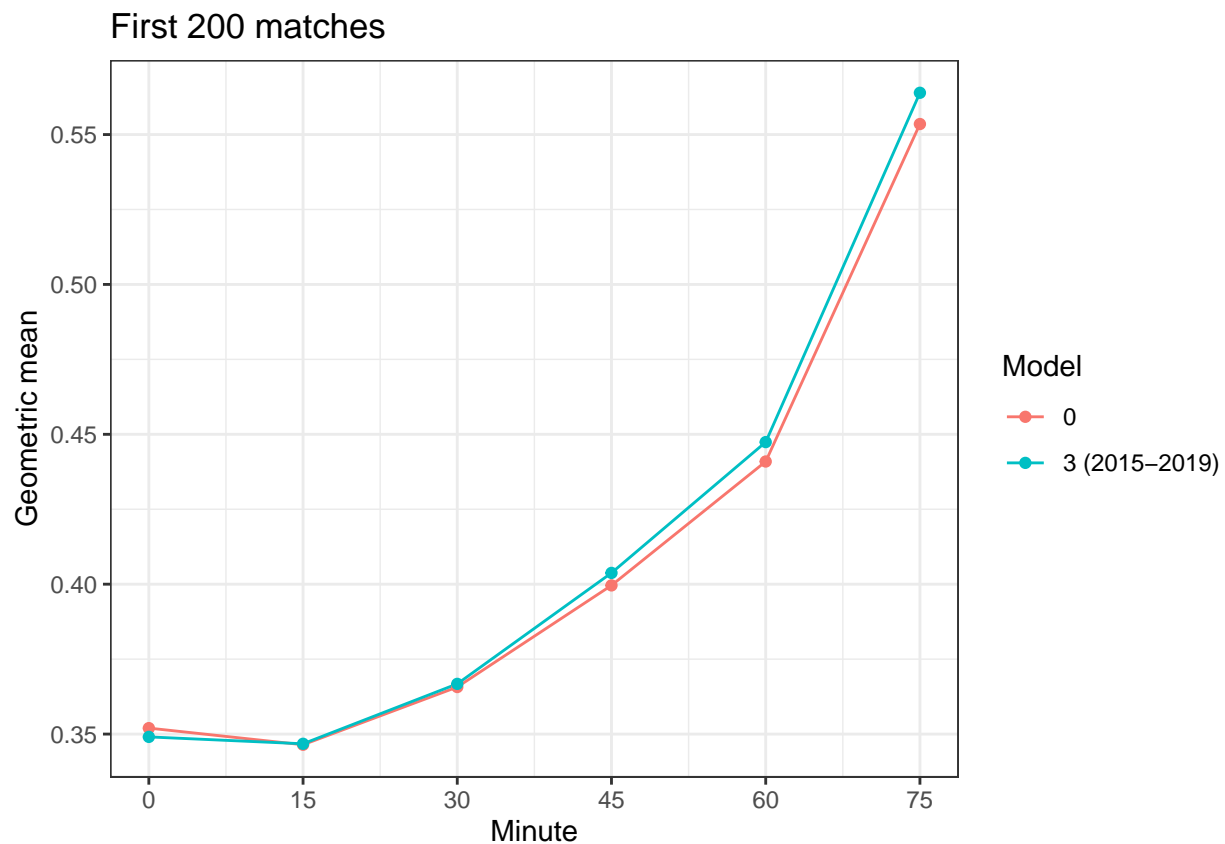
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.3620551	0.3448970	0.3570312	0.3837713	0.4290528	0.5327926
3 (2015-2019)	0.3621657	0.3497791	0.3628308	0.3920926	0.4405226	0.5490799

```
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")
```
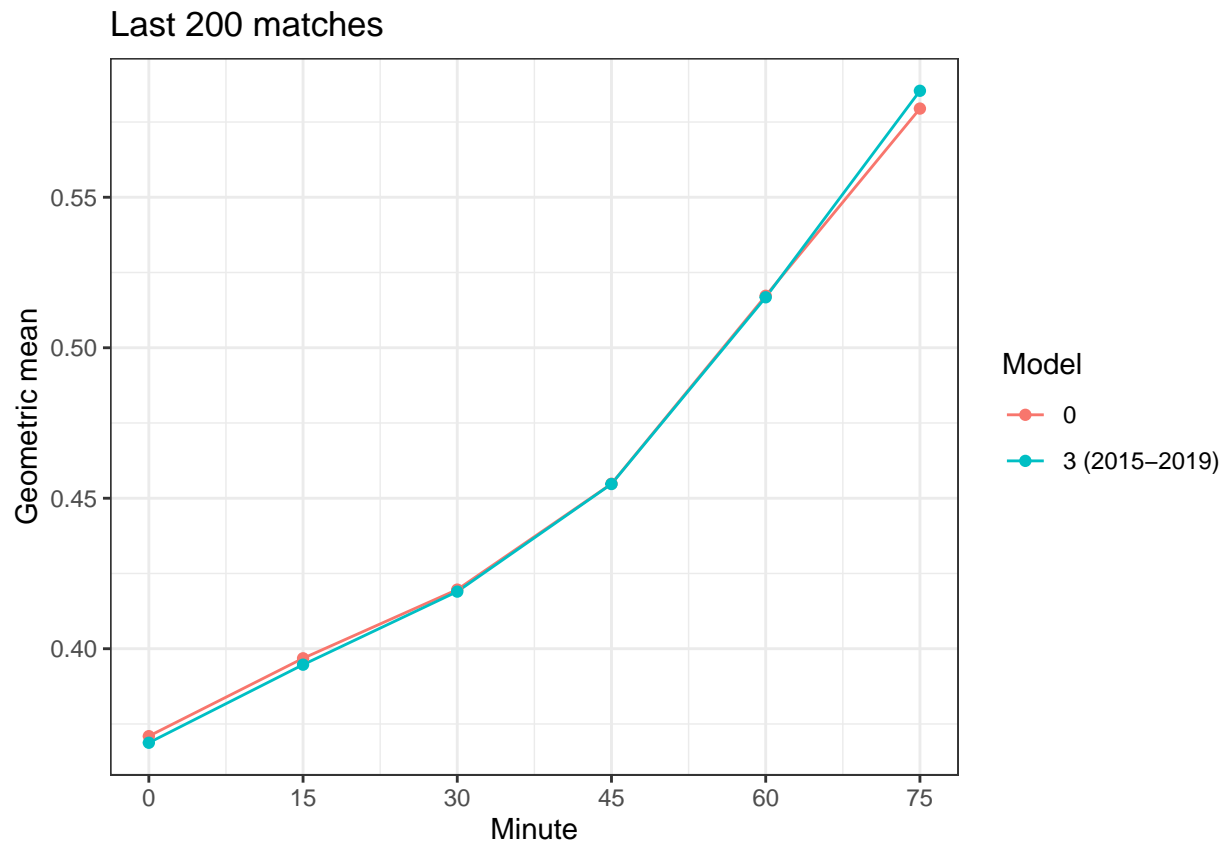


```
first_200 %>%
  pivot_wider(id_cols = "Model", values_from = "RPS", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.3519878	0.3464357	0.3656954	0.3995993	0.4409281	0.5535110
3 (2015-2019)	0.3490592	0.3467571	0.3667497	0.4037598	0.4474074	0.5639305

```
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")
```
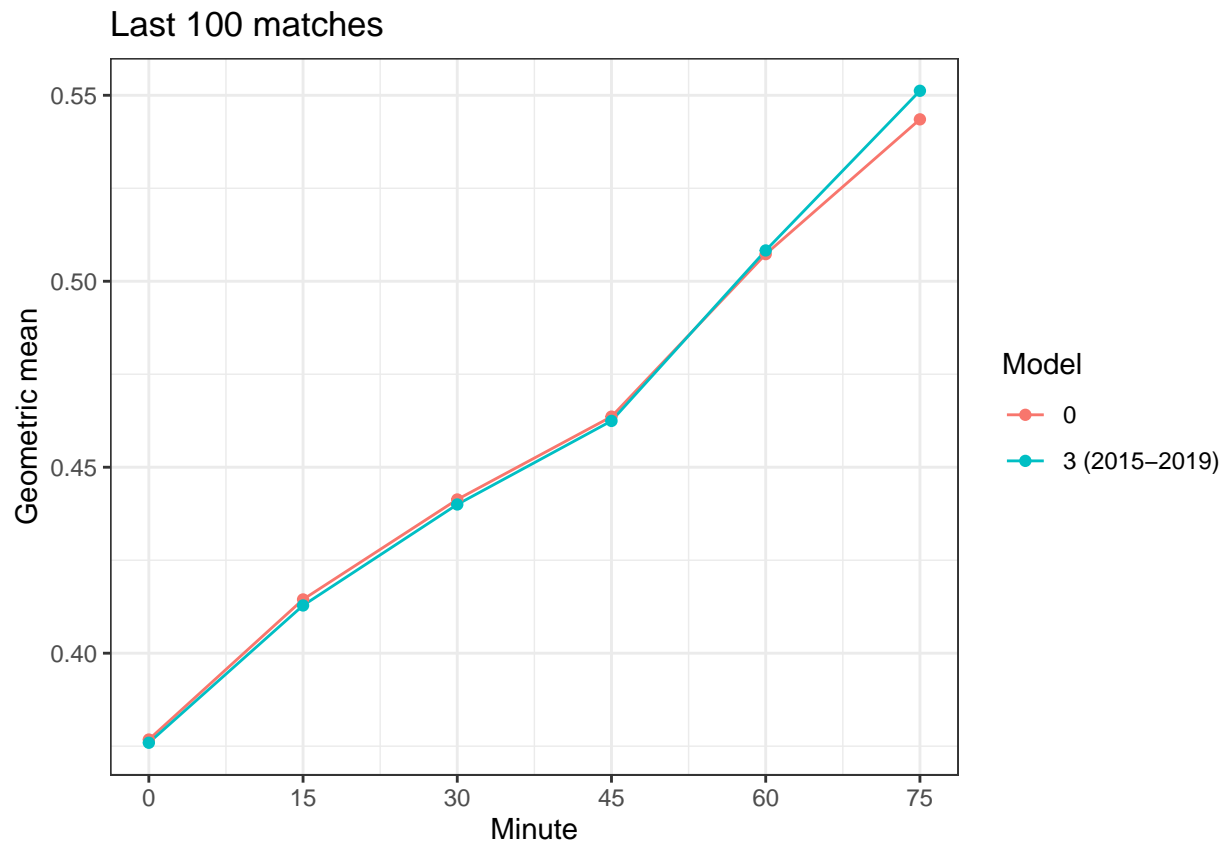


```
last_200 %>%
  pivot_wider(id_cols = "Model", values_from = "RPS", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.3709586	0.3968174	0.4196417	0.4548040	0.5172669	0.5794651
3 (2015-2019)	0.3687581	0.3946998	0.4189798	0.4547298	0.5167623	0.5853762

```
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	0.3767936	0.4144559	0.4413339	0.4636039	0.5072935	0.5435315
3 (2015-2019)	0.3759238	0.4128385	0.4399901	0.4624527	0.5082845	0.5512083

```
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 matches with red cards in the first half
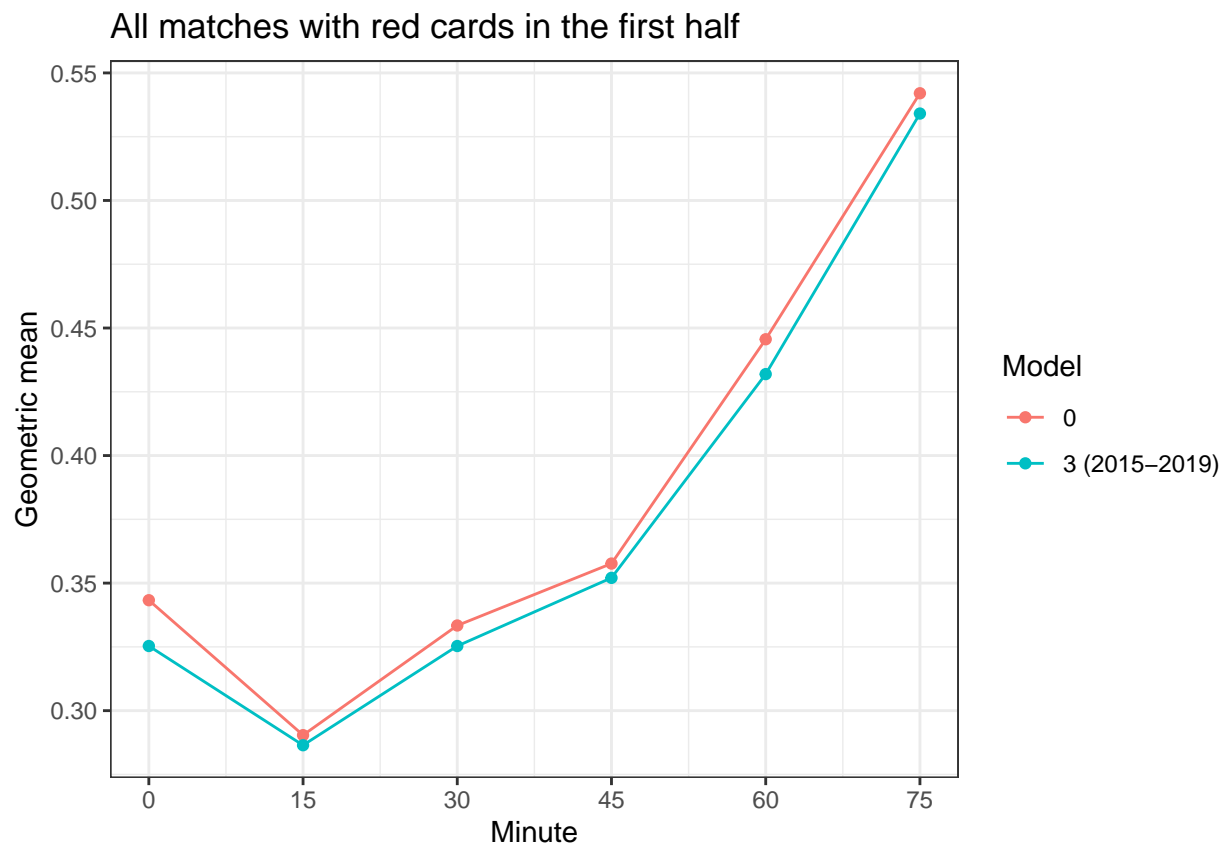
```
all_reds %>%
  pivot_wider(id_cols = "Model", values_from = "RPS", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.3432949	0.2903976	0.3333805	0.3576654	0.445590	0.5420251
3 (2015-2019)	0.3253461	0.2864461	0.3253348	0.3520578	0.431927	0.5340688