Geometric mean

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

load("data/HDA_dc.RData")
load("data/first_matches.RData")

HDA = HDA_dc %>%
  anti_join(first_matches)

nrow(HDA)
```
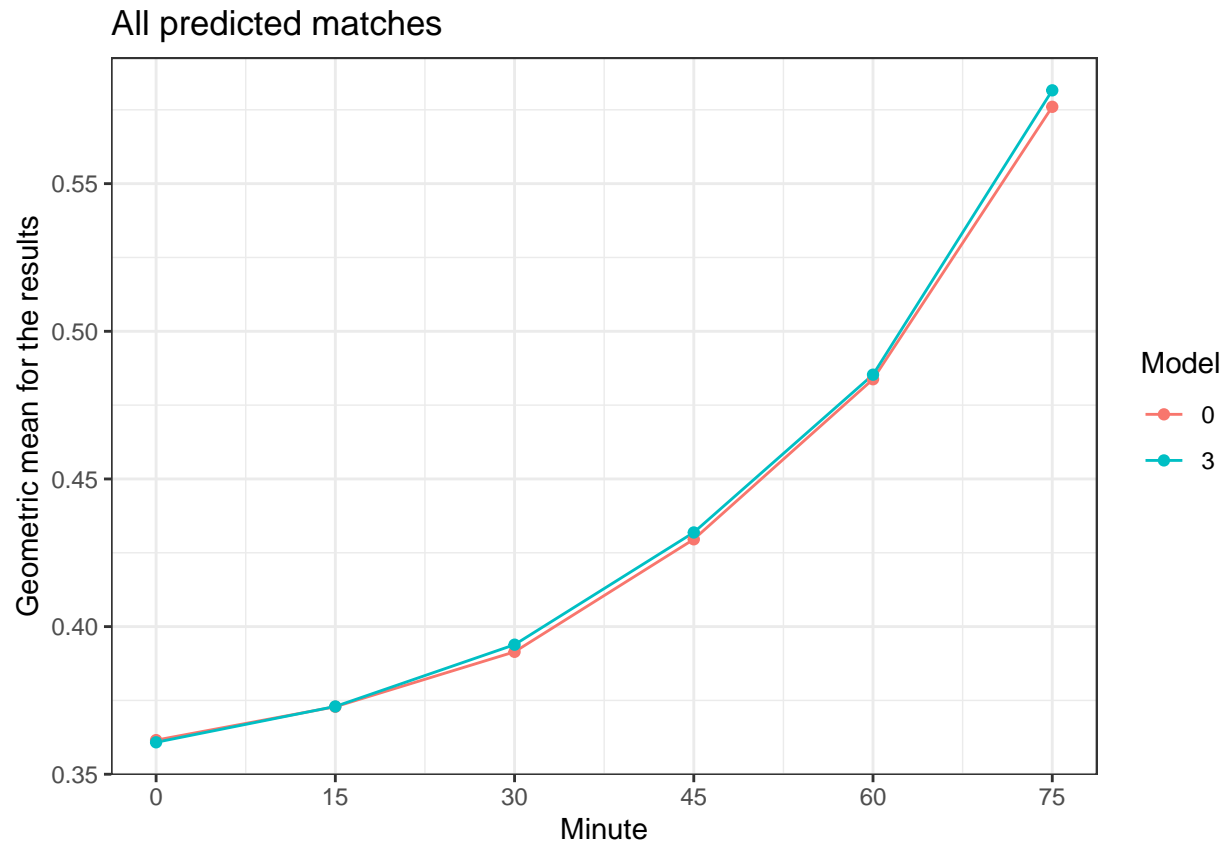
```
## [1] 1858
```

```
results = tibble(GeoMean = apply(HDA[,c(45:56)], 2, EnvStats::geoMean),
                 Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                 Model = factor(c(rep("0", 6),
                                  rep("3", 6))))
results %>%
  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 for the results")
```
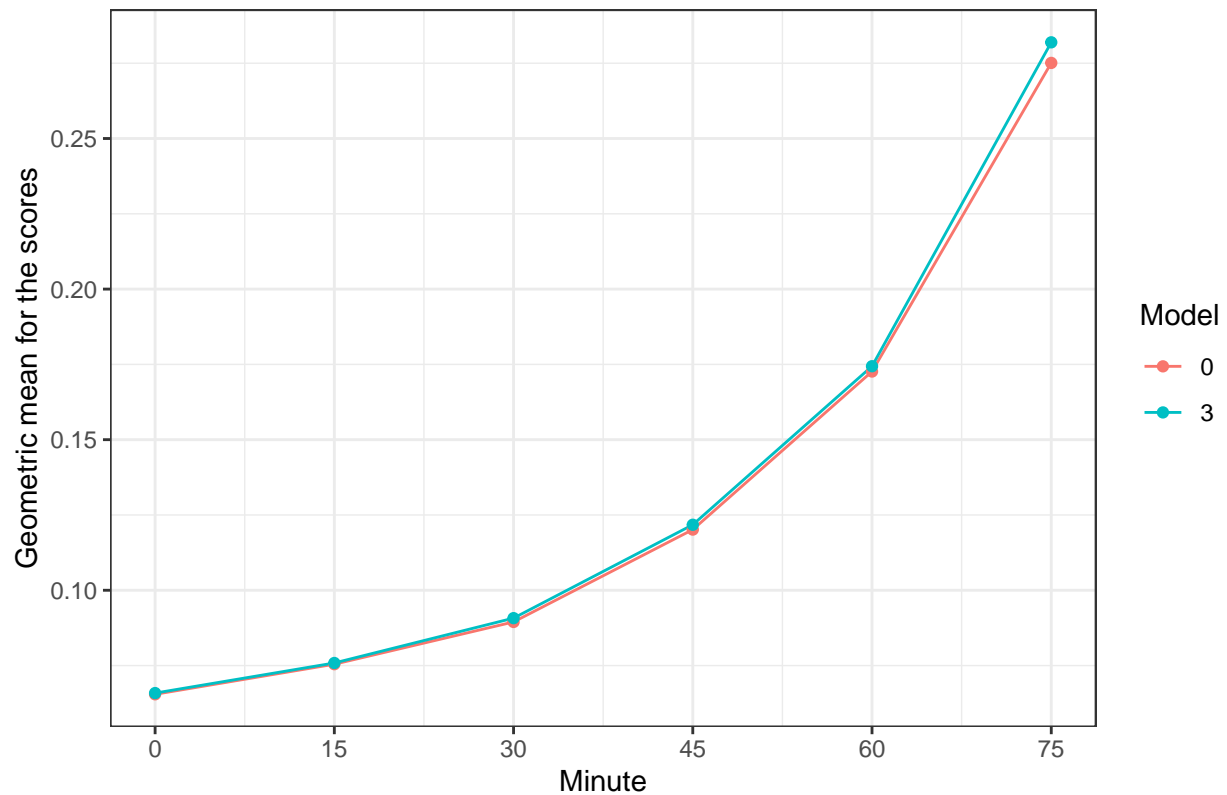
All predicted matches


```
results %>%
  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	0.3615542	0.3728595	0.3914571	0.4295694	0.4837697	0.5760223
3	0.3608178	0.3729549	0.3938743	0.4318778	0.4853132	0.5815989

```
scores = tibble(GeoMean = apply(HDA[,c(57:68)], 2, EnvStats::geoMean),
                Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                Model = factor(c(rep("0", 6),
                                 rep("3", 6))))
scores %>%
  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 for the scores")
```

All predicted matches


```
scores %>%
  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	0.0654583	0.0754529	0.0894830	0.1201492	0.1726362	0.2751038
3	0.0658623	0.0758537	0.0907471	0.1217557	0.1743915	0.2819345

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

matches = reds %>%
  filter(Season > 2015, Half == 1) %>%
  select(Season, Match)

HDA_reds = HDA %>%
  inner_join(matches)
```

```
## Joining, by = c("Season", "Match")
```

```
HDA_no_reds = HDA %>%
  anti_join(matches)
```
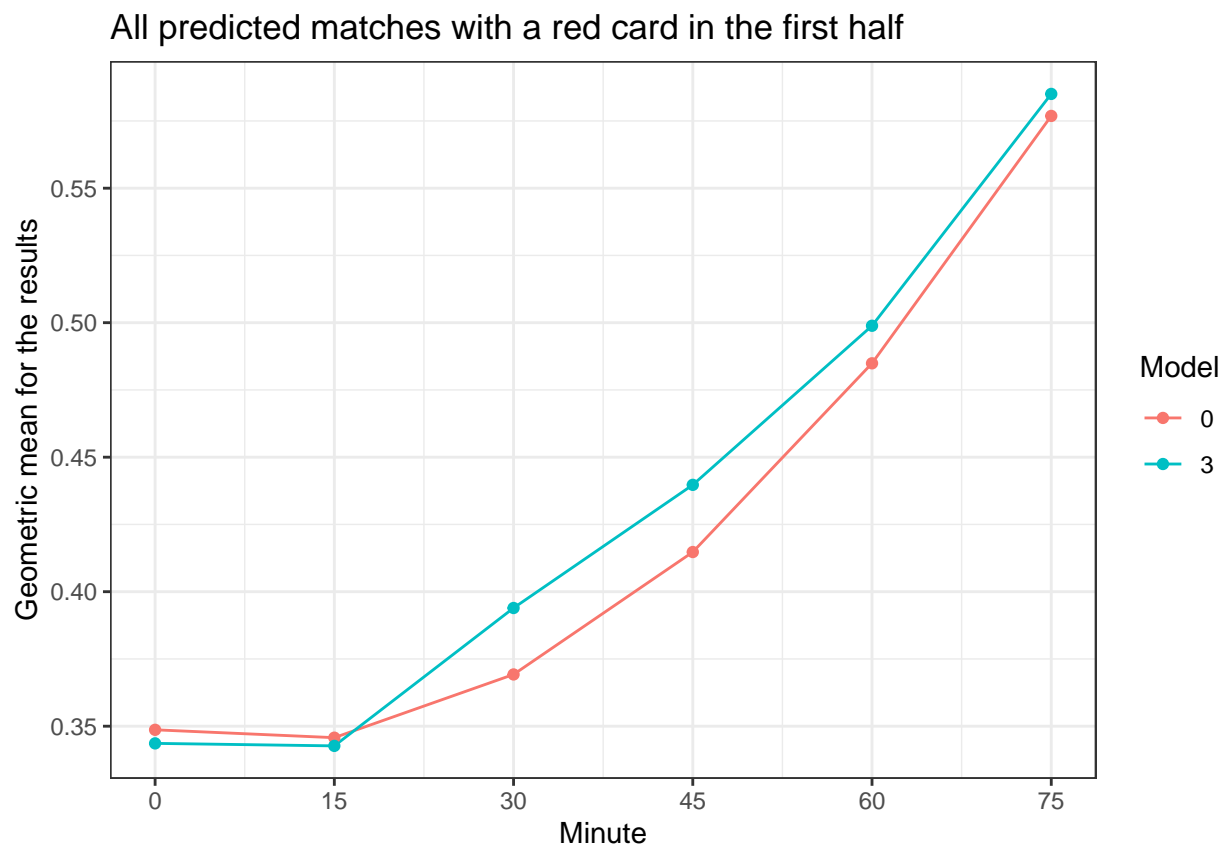
```
## Joining, by = c("Season", "Match")
```

```
nrow(HDA_reds)
```

```
## [1] 82
```

```
results_reds = tibble(GeoMean = apply(HDA_reds[,c(45:56)], 2, EnvStats::geoMean),
                      Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                      Model = factor(c(rep("0", 6),
                                       rep("3", 6))))
results_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 predicted matches with a red card in the first half") +
  ylab("Geometric mean for the results")
```
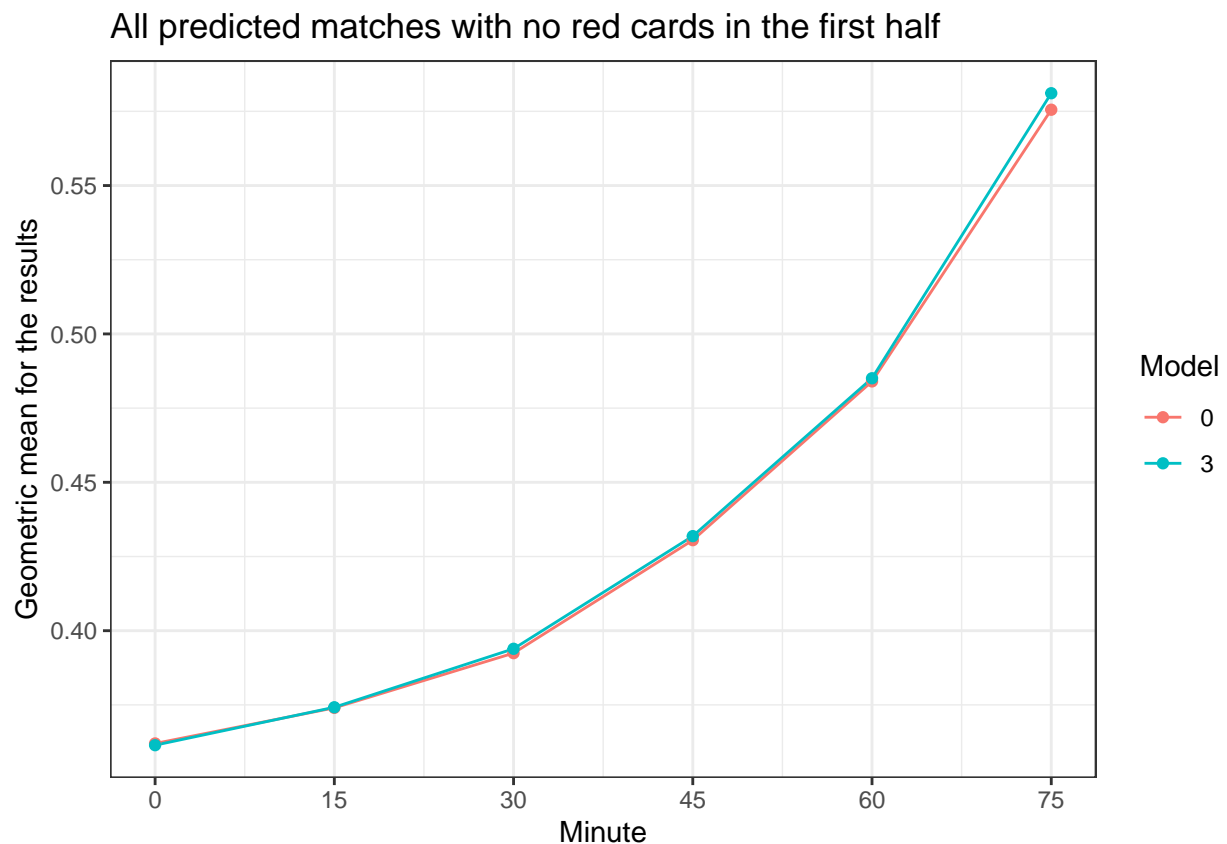
All predicted matches with a red card in the first half


```
results_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	0.3486645	0.3457368	0.3692382	0.4147450	0.4848982	0.5768597
3	0.3436144	0.3426827	0.3939457	0.4397161	0.4988566	0.5850398

```r
results_no_reds = tibble(GeoMean = apply(HDA_no_reds[,c(45:56)], 2, EnvStats::geoMean),
                         Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                         Model = factor(c(rep("0", 6),
                                          rep("3", 6))))
results_no_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 predicted matches with no red cards in the first half") +
  ylab("Geometric mean for the results")
```
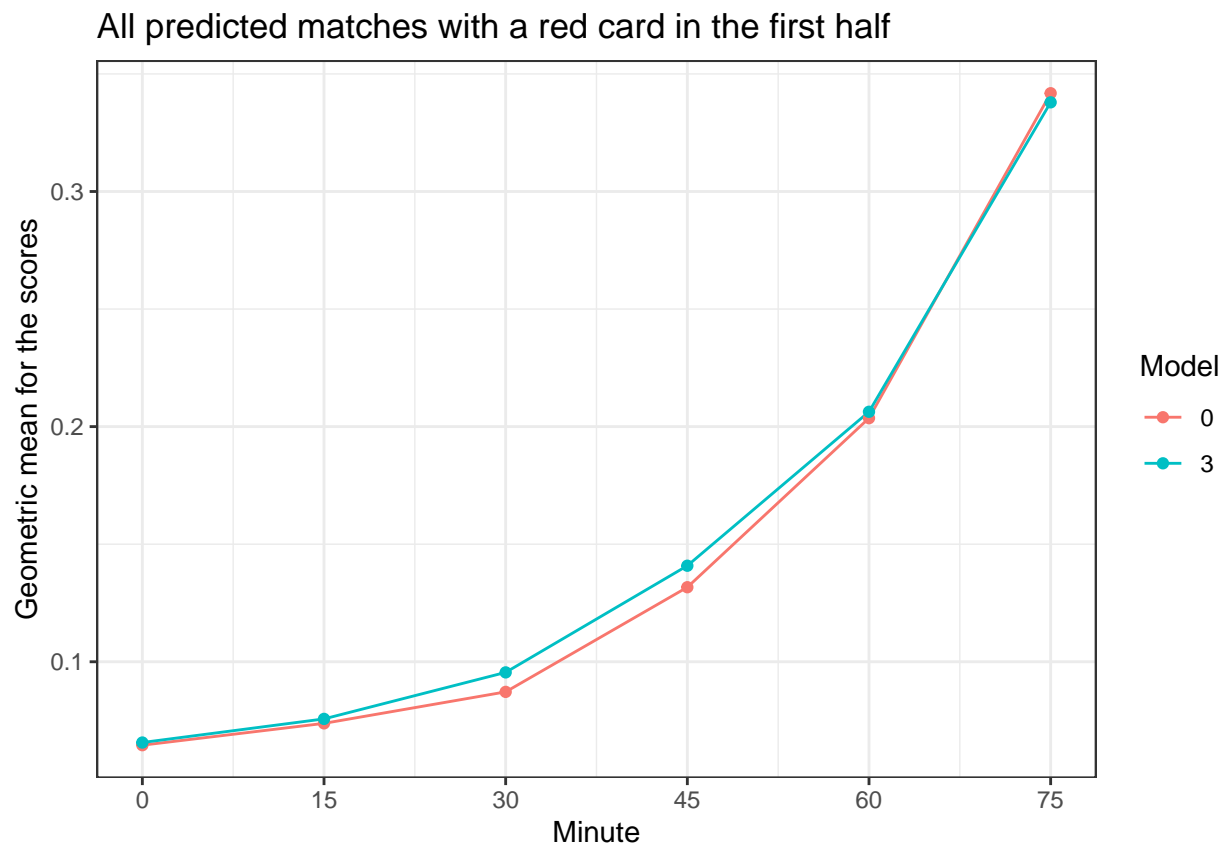

All predicted matches with no red cards in the first half

```r
results_no_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	0.3620003	0.3739905	0.3924370	0.4305123	0.4840060	0.5755491
3	0.3614173	0.3741974	0.3939049	0.4318678	0.4850609	0.5811073

```r
scores_reds = tibble(GeoMean = apply(HDA_reds[,c(57:68)], 2, EnvStats::geoMean),
                     Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                     Model = factor(c(rep("0", 6),
                                      rep("3", 6))))
scores_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 predicted matches with a red card in the first half") +
  ylab("Geometric mean for the scores")
```
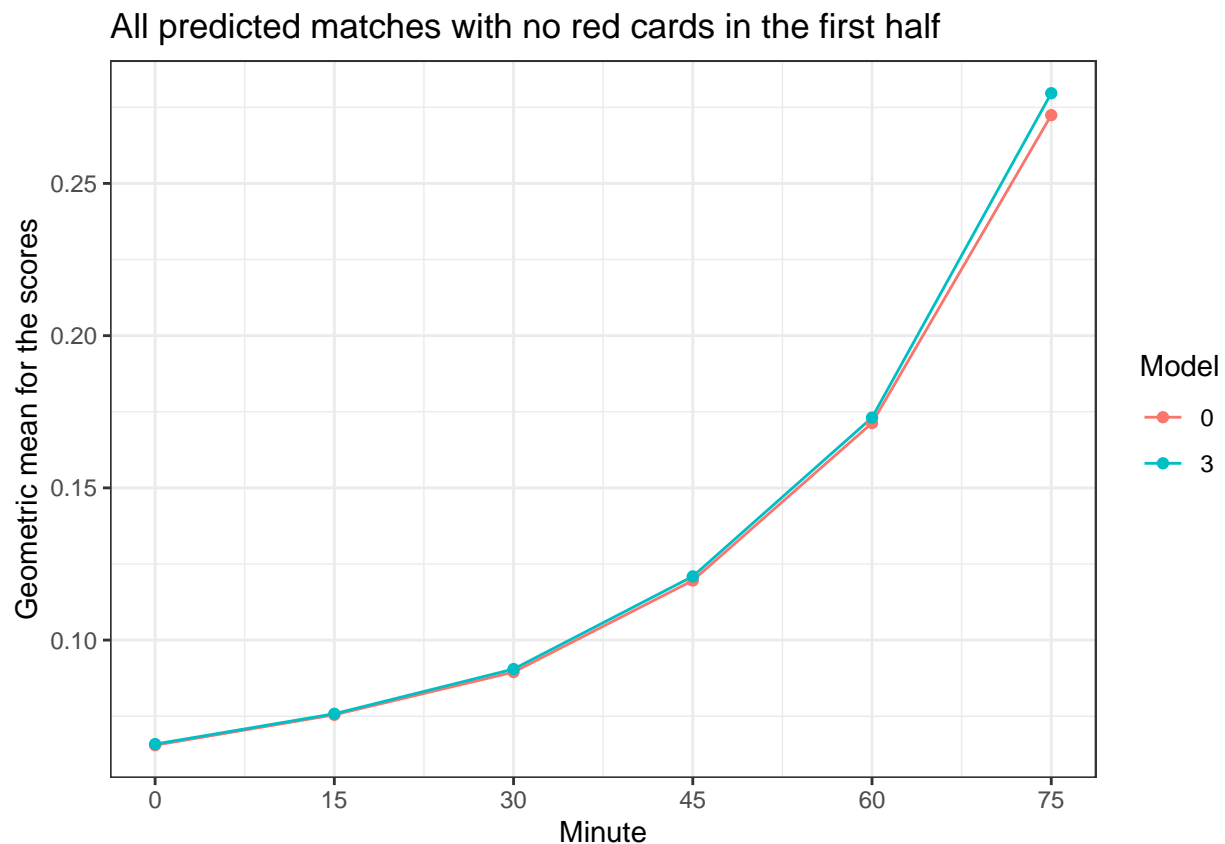

All predicted matches with a red card in the first half

```r
scores_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	0.0645187	0.0738059	0.0871930	0.1316973	0.2036072	0.3417896
3	0.0656550	0.0756983	0.0954845	0.1408456	0.2063144	0.3379100

```r
scores_no_reds = tibble(GeoMean = apply(HDA_no_reds[,c(57:68)], 2, EnvStats::geoMean),
                        Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                        Model = factor(c(rep("0", 6),
                                         rep("3", 6))))
scores_no_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 predicted matches with no red cards in the first half") +
  ylab("Geometric mean for the scores")
```


All predicted matches with no red cards in the first half

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

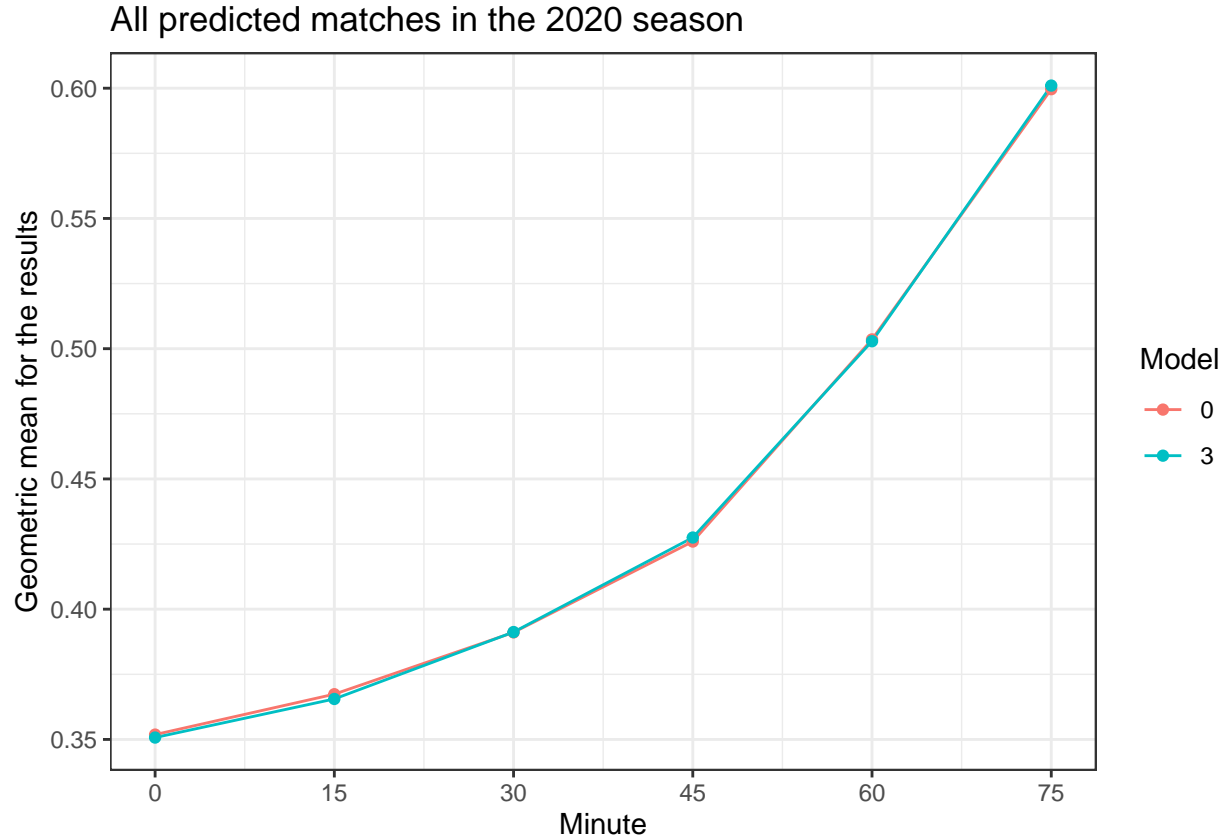| Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.0654788	0.0754585	0.0894830	0.1195775	0.1712945	0.2724196
3	0.0658462	0.0757965	0.0904839	0.1209245	0.1730690	0.2796281

```r
HDA_2020 = HDA %>%
  filter(Season == 2020)

nrow(HDA_2020)
```

```
## [1] 376
```

```r
results_2020 = tibble(GeoMean = apply(HDA_2020[,c(45:56)], 2, EnvStats::geoMean),
                      Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                      Model = factor(c(rep("0", 6),
                                       rep("3", 6))))
results_2020 %>%
  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 in the 2020 season") +
  ylab("Geometric mean for the results")
```
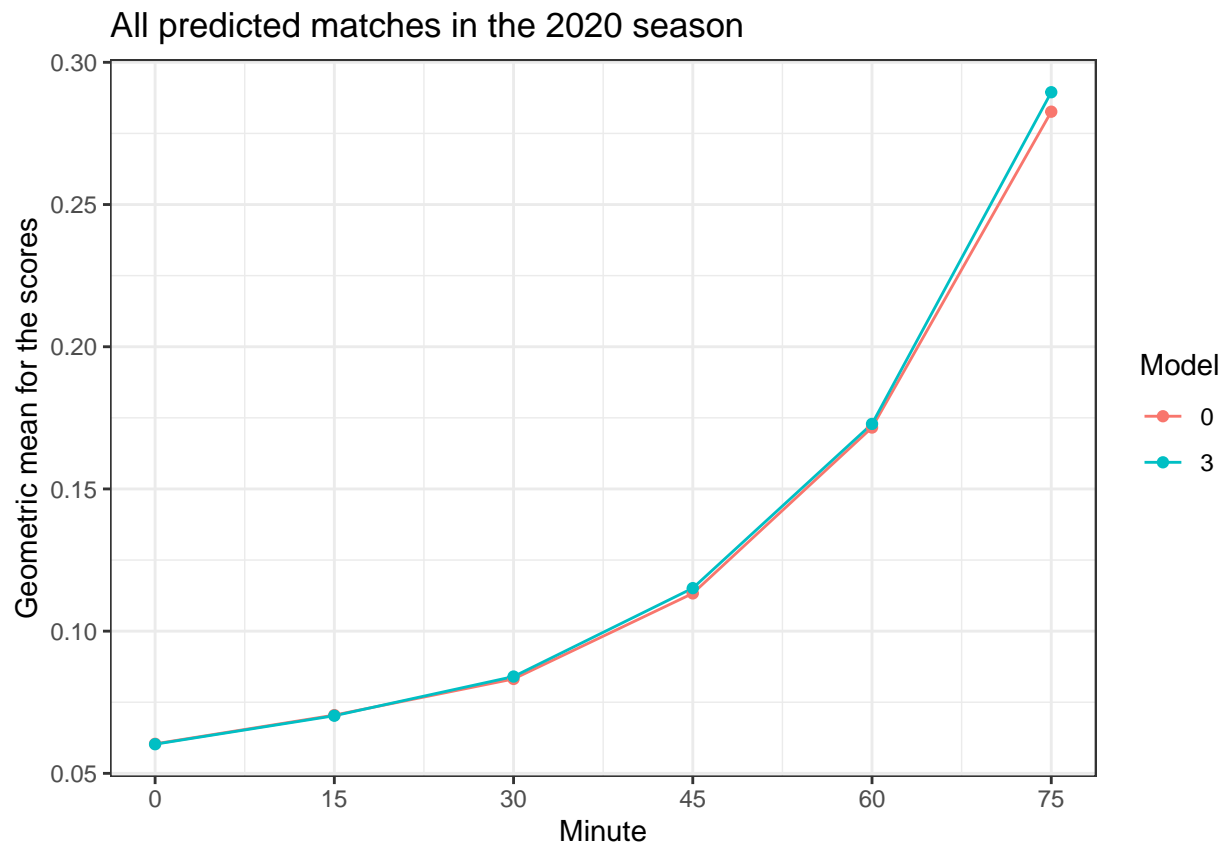

All predicted matches in the 2020 season

```
results_2020 %>%
  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	0.3518546	0.3673683	0.3910833	0.4259717	0.5035654	0.5996244
3	0.3507179	0.3655351	0.3912212	0.4275058	0.5028590	0.6010051

```
scores_2020 = tibble(GeoMean = apply(HDA_2020[,c(57:68)], 2, EnvStats::geoMean),
                     Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 2)),
                     Model = factor(c(rep("0", 6),
                                      rep("3", 6))))
scores_2020 %>%
  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 in the 2020 season") +
  ylab("Geometric mean for the scores")
```

```
scores_2020 %>%
  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	0.0603763	0.0704758	0.0831918	0.1132685	0.1715947	0.2826407
3	0.0602509	0.0702607	0.0840340	0.1151168	0.1728721	0.2894863