Geometric mean

```r
options(OutDec = ",")

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

load("data/HDA_dc_2.RData")
load("data/first_matches.RData")

HDA = HDA_dc %>%
  anti_join(first_matches)

nrow(HDA)
```

```
## [1] 1858
```

```r
HDA[,c(9:158)][which(HDA[,c(9:158)] == 0, arr.ind = TRUE)] = 10^-5
```

```r
results = tibble(GeoMean = apply(HDA[,c(99:128)], 2, EnvStats::geoMean),
                 Minuto = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                 Modelo = factor(c(rep("0", 6),
                                   rep("1", 6),
                                   rep("2", 6),
                                   rep("3", 6),
                                   rep("4", 6)))) %>%
  pivot_wider(names_from = "Modelo", values_from = "GeoMean", names_prefix = "Modelo ") %>%
  mutate(`Modelo 1` = log(`Modelo 1`) - log(`Modelo 0`),
         `Modelo 2` = log(`Modelo 2`) - log(`Modelo 0`),
         `Modelo 3` = log(`Modelo 3`) - log(`Modelo 0`),
         `Modelo 4` = log(`Modelo 4`) - log(`Modelo 0`),
         `Modelo 0` = 0) %>%
  pivot_longer(cols = starts_with("Modelo"), names_to = "Modelo", values_to = "GeoMean") %>%
  mutate(Modelo = stringr::str_replace_all(Modelo, "Modelo ", ""))


p = results %>%
  ggplot(aes(x = Minuto, y = GeoMean, col = Modelo)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  #ggtitle("All predicted matches") +
  ylab("ln(MG(Modelo)) - ln(MG(Modelo 0))")
p
```
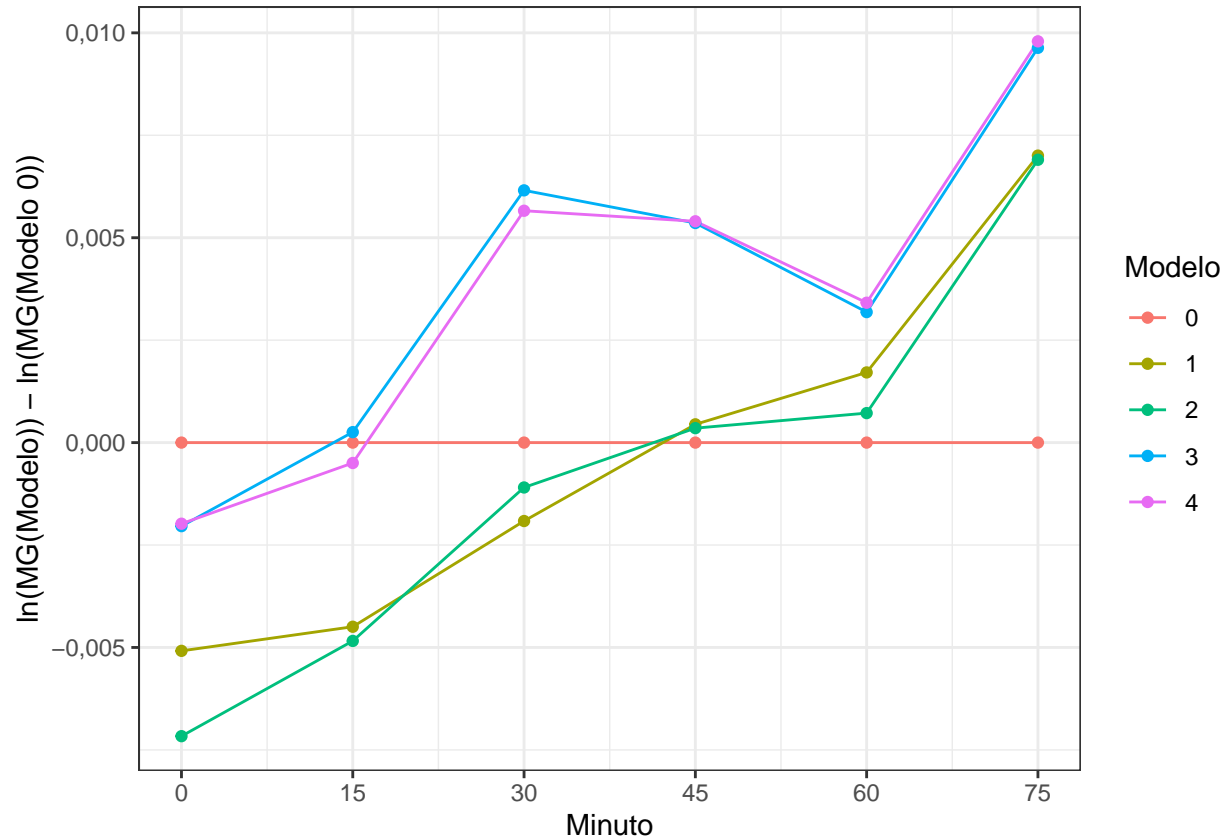
```
ggsave(filename = paste0("plots/results.png"),
       plot = p, width = 10, height = 5, dpi = 1000)

results %>%
  pivot_wider(id_cols = "Modelo", values_from = "GeoMean", names_from = "Minuto",
              names_prefix = "Minuto ") %>%
  kable()
```

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
0	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000
1	-0,0050830	-0,0044942	-0,0019138	0,0004458	0,0017122	0,0070038
2	-0,0071644	-0,0048408	-0,0010950	0,0003527	0,0007197	0,0069024
3	-0,0020390	0,0002557	0,0061560	0,0053594	0,0031854	0,0096346
4	-0,0019799	-0,0004976	0,0056584	0,0054008	0,0034135	0,0097926

```
scores = tibble(GeoMean = apply(HDA[,c(129:158)], 2, EnvStats::geoMean),
                Minuto = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                Modelo = factor(c(rep("0", 6),
                                  rep("1", 6),
                                  rep("2", 6),
                                  rep("3", 6),
                                  rep("4", 6)))) %>%
  pivot_wider(names_from = "Modelo", values_from = "GeoMean", names_prefix = "Modelo ") %>%
```

```r
  mutate(`Modelo 1` = log(`Modelo 1`) - log(`Modelo 0`),
         `Modelo 2` = log(`Modelo 2`) - log(`Modelo 0`),
         `Modelo 3` = log(`Modelo 3`) - log(`Modelo 0`),
         `Modelo 4` = log(`Modelo 4`) - log(`Modelo 0`),
         `Modelo 0` = 0) %>%
  pivot_longer(cols = starts_with("Modelo"), names_to = "Modelo", values_to = "GeoMean") %>%
  mutate(Modelo = stringr::str_replace_all(Modelo, "Modelo ", ""))

p = scores %>%
  ggplot(aes(x = Minuto, y = GeoMean, col = Modelo)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  #ggtitle("All predicted matches") +
  ylab("ln(MG(Modelo)) - ln(MG(Modelo 0))")
p
```
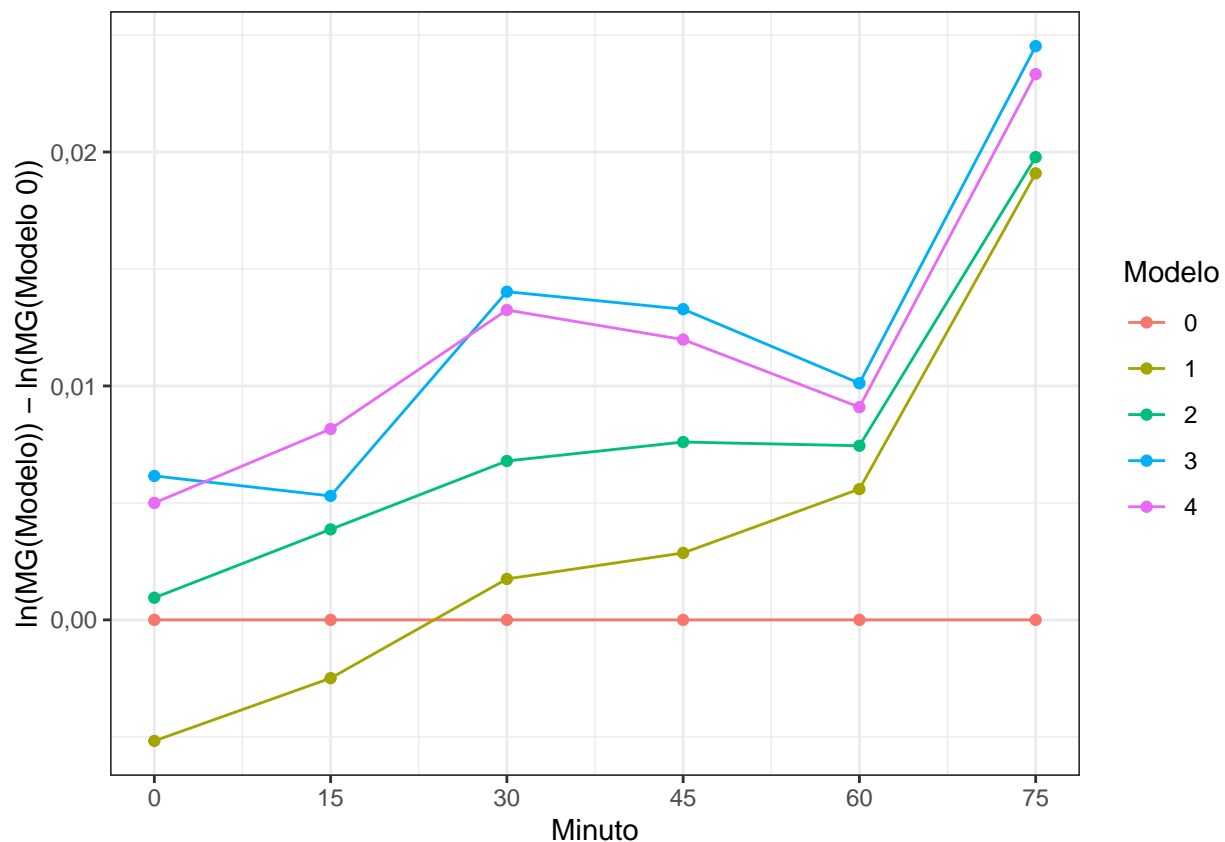


```r
ggsave(filename = paste0("plots/scores.png"),
       plot = p, width = 10, height = 5, dpi = 1000)

scores %>%
  pivot_wider(id_cols = "Modelo", values_from = "GeoMean", names_from = "Minuto",
              names_prefix = "Minuto ") %>%
  kable()
```

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
0	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000
1	-0,0051754	-0,0024916	0,0017509	0,0028609	0,0055911	0,0190873
2	0,0009452	0,0038688	0,0067928	0,0076022	0,0074422	0,0197810
3	0,0061534	0,0052974	0,0140277	0,0132818	0,0101162	0,0245261
4	0,0050004	0,0081606	0,0132462	0,0119814	0,0090928	0,0233283

```r
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(99:128)], 2, EnvStats::geoMean),
                      Minuto = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                      Modelo = factor(c(rep("0", 6),
                                        rep("1", 6),
                                        rep("2", 6),
                                        rep("3", 6),
                                        rep("4", 6)))) %>%
  pivot_wider(names_from = "Modelo", values_from = "GeoMean", names_prefix = "Modelo ") %>%
  mutate(`Modelo 1` = log(`Modelo 1`) - log(`Modelo 0`),
         `Modelo 2` = log(`Modelo 2`) - log(`Modelo 0`),
         `Modelo 3` = log(`Modelo 3`) - log(`Modelo 0`),
         `Modelo 4` = log(`Modelo 4`) - log(`Modelo 0`),
         `Modelo 0` = 0) %>%
  pivot_longer(cols = starts_with("Modelo"), names_to = "Modelo", values_to = "GeoMean") %>%
  mutate(Modelo = stringr::str_replace_all(Modelo, "Modelo ", ""))

p = results_reds %>%
  ggplot(aes(x = Minuto, y = GeoMean, col = Modelo)) +
  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("ln(MG(Modelo)) - ln(MG(Modelo 0))")
p
```
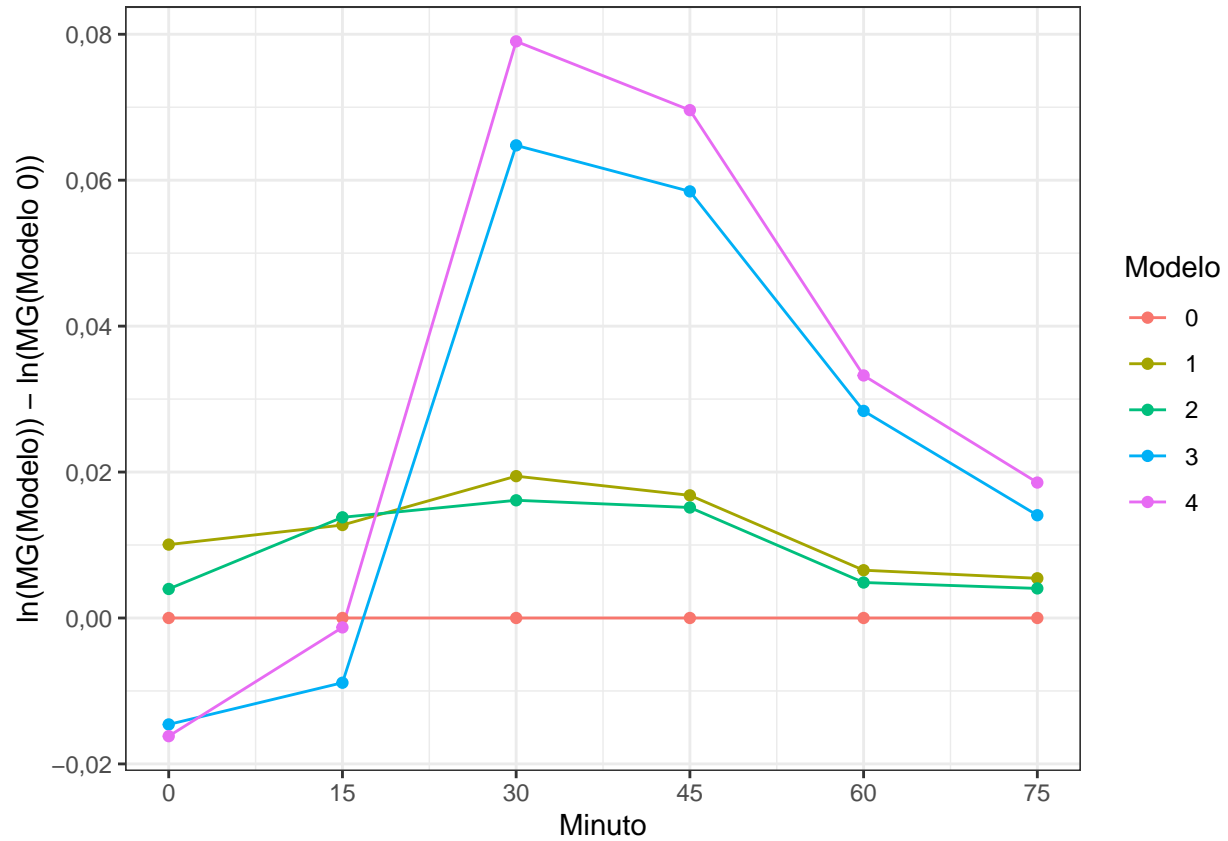
```
ggsave(filename = paste0("plots/results_reds.png"),
       plot = p, width = 10, height = 5, dpi = 1000)

results_reds %>%
  pivot_wider(id_cols = "Modelo", values_from = "GeoMean", names_from = "Minuto",
              names_prefix = "Minuto ") %>%
  kable()
```

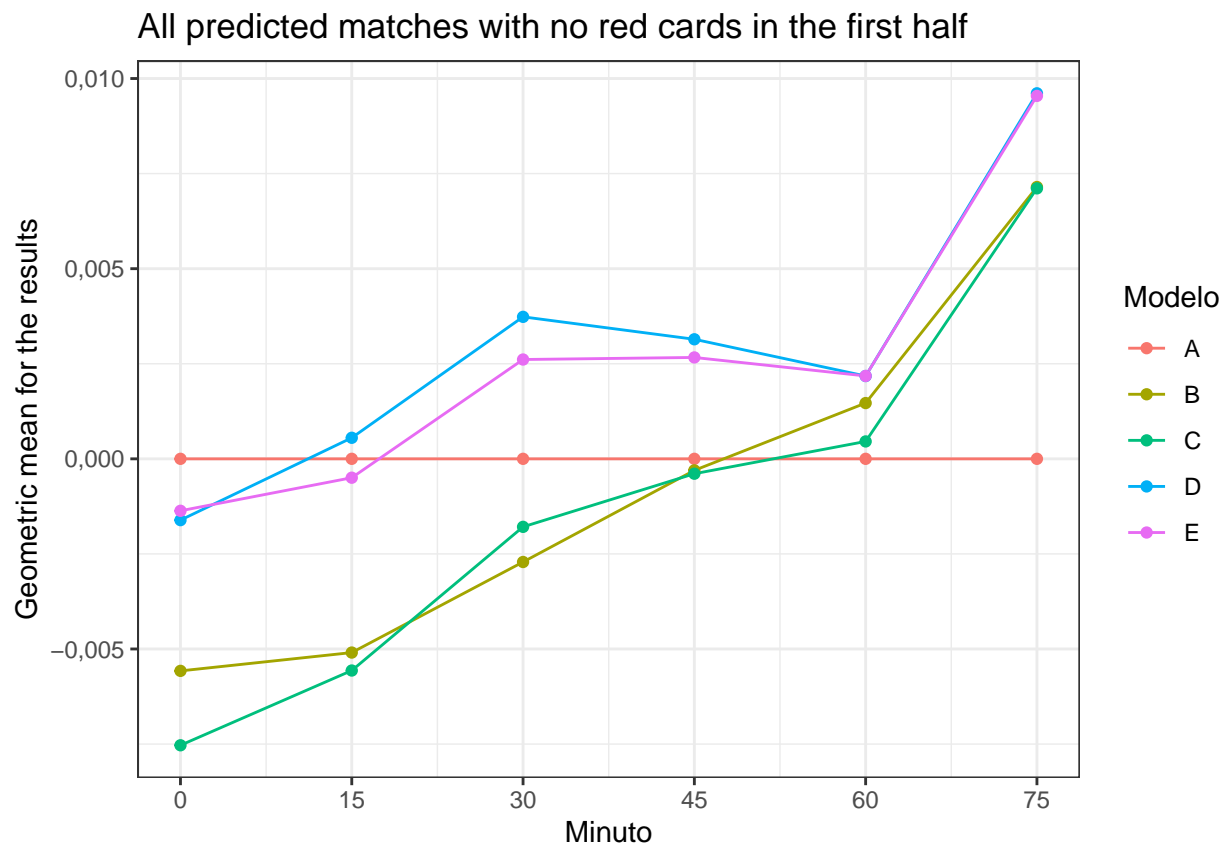| Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
0	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000
1	0,0100593	0,0127398	0,0194376	0,0168047	0,0065459	0,0054407
2	0,0039783	0,0137975	0,0161298	0,0151433	0,0048625	0,0040526
3	-0,0145901	-0,0088729	0,0647713	0,0584654	0,0283797	0,0140808
4	-0,0161959	-0,0012889	0,0790279	0,0696009	0,0332385	0,0185605

```
results_no_reds = tibble(GeoMean = apply(HDA_no_reds[,c(99:128)], 2, EnvStats::geoMean),
                         Minuto = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                         Modelo = factor(c(rep("A", 6),
                                           rep("B", 6),
                                           rep("C", 6),
                                           rep("D", 6),
                                           rep("E", 6)))) %>%
  pivot_wider(names_from = "Modelo", values_from = "GeoMean", names_prefix = "Modelo ") %>%
```

```
  mutate(`Modelo B` = log(`Modelo B`) - log(`Modelo A`),
         `Modelo C` = log(`Modelo C`) - log(`Modelo A`),
         `Modelo D` = log(`Modelo D`) - log(`Modelo A`),
         `Modelo E` = log(`Modelo E`) - log(`Modelo A`),
         `Modelo A` = 0) %>%
  pivot_longer(cols = starts_with("Modelo"), names_to = "Modelo", values_to = "GeoMean") %>%
  mutate(Modelo = stringr::str_replace_all(Modelo, "Modelo ", ""))

results_no_reds %>%
  ggplot(aes(x = Minuto, y = GeoMean, col = Modelo)) +
  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")
```
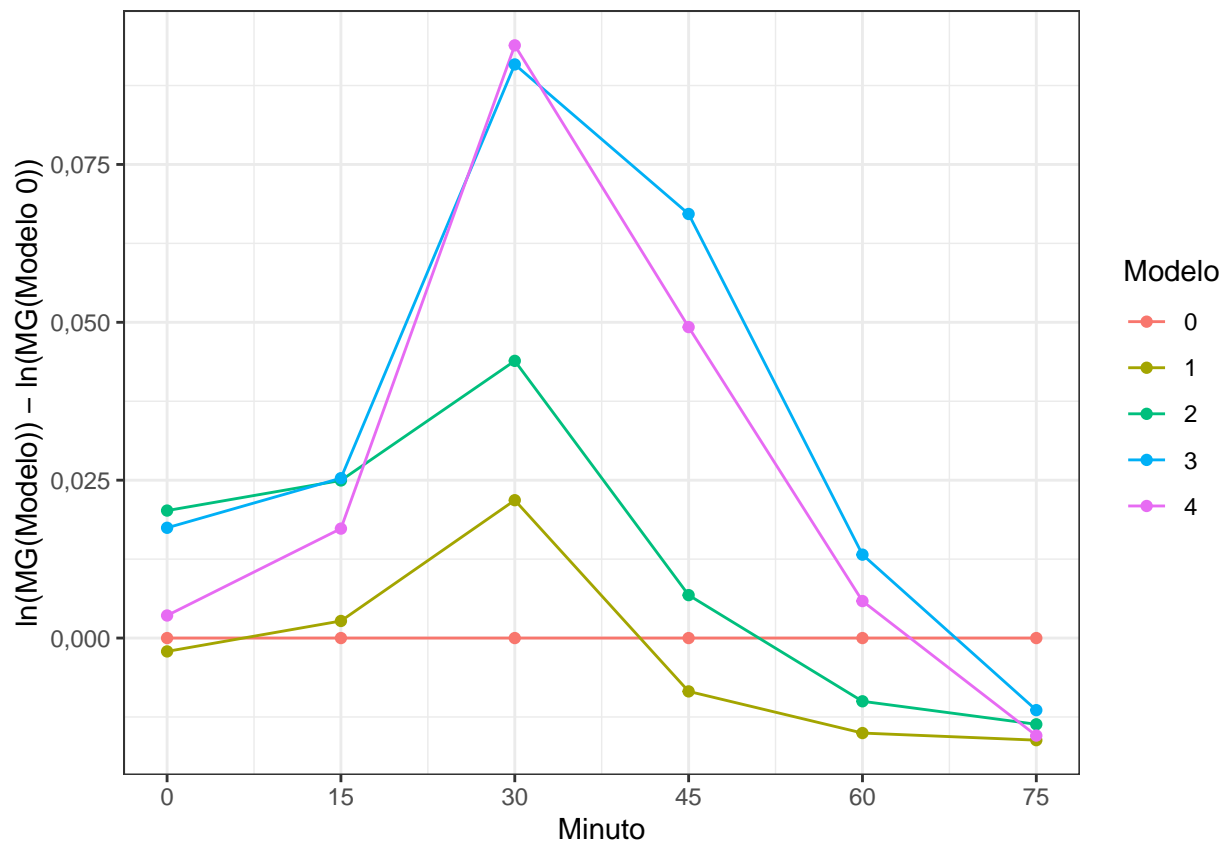


```
results_no_reds %>%
  pivot_wider(id_cols = "Modelo", values_from = "GeoMean", names_from = "Minuto",
              names_prefix = "Minuto ") %>%
  kable()
```

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
A	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000
B	-0,0055757	-0,0050940	-0,0027127	-0,0003030	0,0014663	0,0071475
C	-0,0075320	-0,0055662	-0,0017883	-0,0003913	0,0004570	0,0071122
D	-0,0016118	0,0005530	0,0037334	0,0031436	0,0021773	0,0096109
E	-0,0013664	-0,0004966	0,0026115	0,0026658	0,0021795	0,0095432

```r
scores_reds = tibble(GeoMean = apply(HDA_reds[,c(129:158)], 2, EnvStats::geoMean),
                     Minuto = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                     Modelo = factor(c(rep("0", 6),
                                       rep("1", 6),
                                       rep("2", 6),
                                       rep("3", 6),
                                       rep("4", 6)))) %>%
  pivot_wider(names_from = "Modelo", values_from = "GeoMean", names_prefix = "Modelo ") %>%
  mutate(`Modelo 1` = log(`Modelo 1`) - log(`Modelo 0`),
         `Modelo 2` = log(`Modelo 2`) - log(`Modelo 0`),
         `Modelo 3` = log(`Modelo 3`) - log(`Modelo 0`),
         `Modelo 4` = log(`Modelo 4`) - log(`Modelo 0`),
         `Modelo 0` = 0) %>%
  pivot_longer(cols = starts_with("Modelo"), names_to = "Modelo", values_to = "GeoMean") %>%
  mutate(Modelo = stringr::str_replace_all(Modelo, "Modelo ", ""))

p = scores_reds %>%
  ggplot(aes(x = Minuto, y = GeoMean, col = Modelo)) +
  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("ln(MG(Modelo)) - ln(MG(Modelo 0))")
p
```

```
ggsave(filename = paste0("plots/scores_reds.png"),
       plot = p, width = 10, height = 5, dpi = 1000)

scores_reds %>%
  pivot_wider(id_cols = "Modelo", values_from = "GeoMean", names_from = "Minuto",
              names_prefix = "Minuto ") %>%
  kable()
```

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
0	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000
1	-0,0021167	0,0027039	0,0218298	-0,0084463	-0,0150459	-0,0161702
2	0,0201882	0,0249565	0,0438951	0,0067985	-0,0100154	-0,0136582
3	0,0174591	0,0253175	0,0908397	0,0671582	0,0132089	-0,0114157
4	0,0035712	0,0173301	0,0938627	0,0492465	0,0058546	-0,0154456

```
scores_no_reds = tibble(GeoMean = apply(HDA_no_reds[,c(129:158)], 2, EnvStats::geoMean),
                        Minuto = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                        Modelo = factor(c(rep("A", 6),
                                          rep("B", 6),
                                          rep("C", 6),
                                          rep("D", 6),
                                          rep("E", 6)))) %>%
  pivot_wider(names_from = "Modelo", values_from = "GeoMean", names_prefix = "Modelo ") %>%
```

```
  mutate(`Modelo B` = log(`Modelo B`) - log(`Modelo A`),
         `Modelo C` = log(`Modelo C`) - log(`Modelo A`),
         `Modelo D` = log(`Modelo D`) - log(`Modelo A`),
         `Modelo E` = log(`Modelo E`) - log(`Modelo A`),
         `Modelo A` = 0) %>%
  pivot_longer(cols = starts_with("Modelo"), names_to = "Modelo", values_to = "GeoMean") %>%
  mutate(Modelo = stringr::str_replace_all(Modelo, "Modelo ", ""))

scores_no_reds %>%
  ggplot(aes(x = Minuto, y = GeoMean, col = Modelo)) +
  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
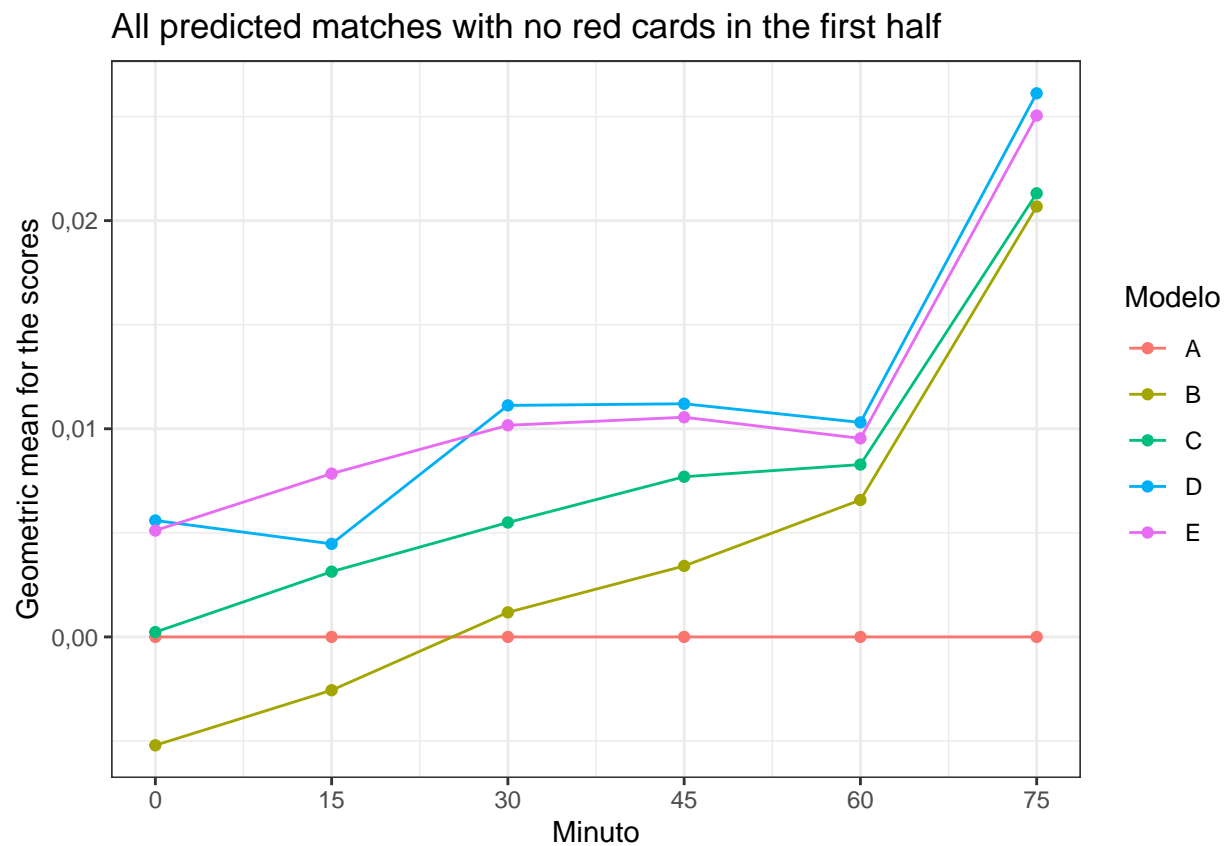
```
scores_no_reds %>%
  pivot_wider(id_cols = "Modelo", values_from = "GeoMean", names_from = "Minuto",
              names_prefix = "Minuto ") %>%
  kable()
```

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
A	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000
B	-0,0052060	-0,0025607	0,0011794	0,0034094	0,0065777	0,0206790
C	0,0002307	0,0031352	0,0054965	0,0076958	0,0082800	0,0213127
D	0,0055947	0,0044703	0,0111232	0,0112016	0,0103060	0,0261170
E	0,0051118	0,0078454	0,0101676	0,0105542	0,0095401	0,0250477

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

nrow(HDA_2020)
```

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

```r
results_2020 = tibble(GeoMean = apply(HDA_2020[,c(99:128)], 2, EnvStats::geoMean),
                      Minuto = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                      Modelo = factor(c(rep("A", 6),
                                        rep("B", 6),
                                        rep("C", 6),
                                        rep("D", 6),
                                        rep("E", 6)))) %>%
  pivot_wider(names_from = "Modelo", values_from = "GeoMean", names_prefix = "Modelo ") %>%
  mutate(`Modelo B` = log(`Modelo B`) - log(`Modelo A`),
         `Modelo C` = log(`Modelo C`) - log(`Modelo A`),
         `Modelo D` = log(`Modelo D`) - log(`Modelo A`),
         `Modelo E` = log(`Modelo E`) - log(`Modelo A`),
         `Modelo A` = 0) %>%
  pivot_longer(cols = starts_with("Modelo"), names_to = "Modelo", values_to = "GeoMean") %>%
  mutate(Modelo = stringr::str_replace_all(Modelo, "Modelo ", ""))

results_2020 %>%
  ggplot(aes(x = Minuto, y = GeoMean, col = Modelo)) +
  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
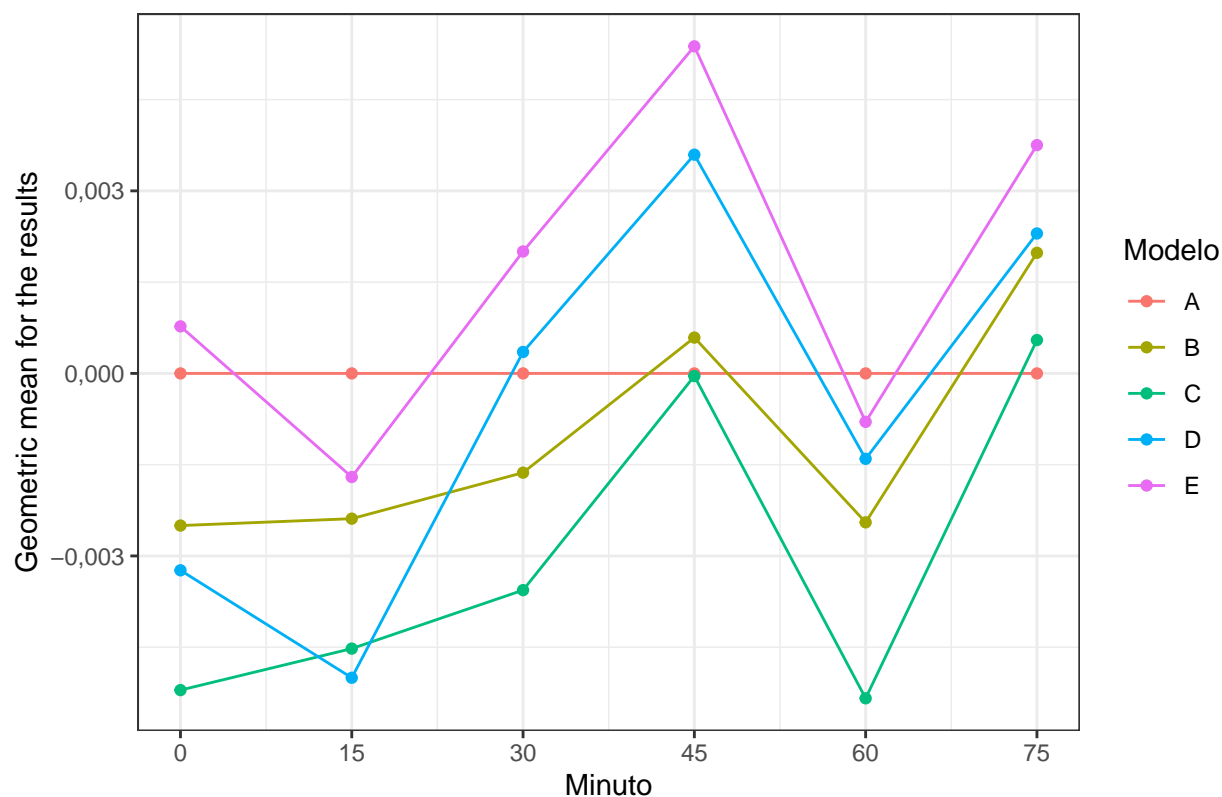

```
results_2020 %>%
  pivot_wider(id_cols = "Modelo", values_from = "GeoMean", names_from = "Minuto",
              names_prefix = "Minuto ") %>%
  kable()
```

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
A	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000
B	-0,0025007	-0,0023876	-0,0016313	0,0005890	-0,0024479	0,0019818
C	-0,0052050	-0,0045233	-0,0035622	-0,0000434	-0,0053399	0,0005480
D	-0,0032359	-0,0050027	0,0003524	0,0035951	-0,0014039	0,0023000
E	0,0007733	-0,0017027	0,0020043	0,0053758	-0,0007964	0,0037524

```
scores_2020 = tibble(GeoMean = apply(HDA_2020[,c(129:158)], 2, EnvStats::geoMean),
                     Minuto = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                     Modelo = factor(c(rep("A", 6),
                                       rep("B", 6),
                                       rep("C", 6),
                                       rep("D", 6),
                                       rep("E", 6)))) %>%
  pivot_wider(names_from = "Modelo", values_from = "GeoMean", names_prefix = "Modelo ") %>%
  mutate(`Modelo B` = log(`Modelo B`) - log(`Modelo A`),
         `Modelo C` = log(`Modelo C`) - log(`Modelo A`),
         `Modelo D` = log(`Modelo D`) - log(`Modelo A`),
```

```
        `Modelo E` = log(`Modelo E`) - log(`Modelo A`),
        `Modelo A` = 0) %>%
  pivot_longer(cols = starts_with("Modelo"), names_to = "Modelo", values_to = "GeoMean") %>%
  mutate(Modelo = stringr::str_replace_all(Modelo, "Modelo ", ""))

scores_2020 %>%
  ggplot(aes(x = Minuto, y = GeoMean, col = Modelo)) +
  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")
```
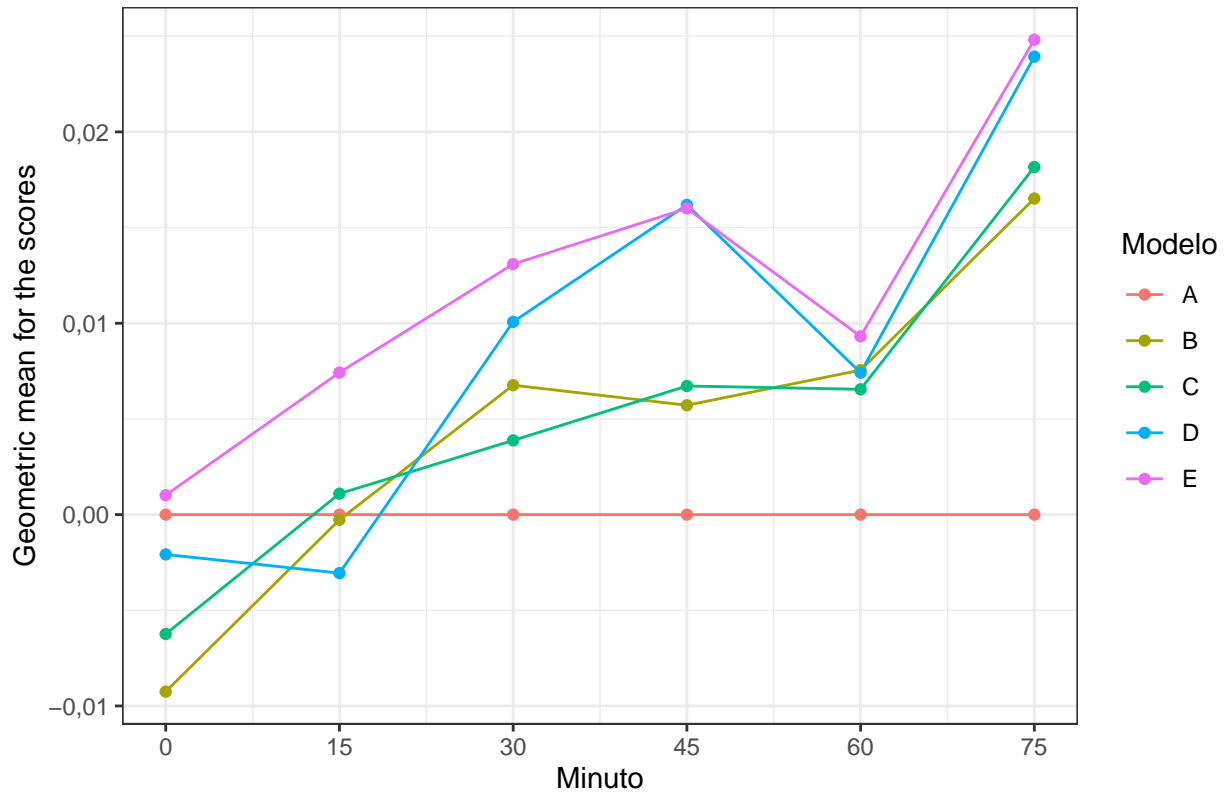


```
scores_2020 %>%
  pivot_wider(id_cols = "Modelo", values_from = "GeoMean", names_from = "Minuto",
              names_prefix = "Minuto ") %>%
  kable()
```

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
A	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000
B	-0,0092506	-0,0002693	0,0067635	0,0057149	0,0075514	0,0165179
C	-0,0062386	0,0010945	0,0038794	0,0067212	0,0065486	0,0181650
D	-0,0020800	-0,0030570	0,0100732	0,0161854	0,0074168	0,0239314

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
E	0,0010146	0,0074241	0,0130901	0,0160045	0,0093150	0,0248199

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

```r
at_45 = results %>%
  select(Season, Match) %>%
  filter(Season > 2015)
```

```r
home_score_at_45 <- function(season, match) {
  goals %>%
    filter(Season == season,
           Match == match,
           Team == 1,
           Half == 1) %>%
    nrow()
}

away_score_at_45 <- function(season, match) {
  goals %>%
    filter(Season == season,
           Match == match,
           Team == 2,
           Half == 1) %>%
    nrow()
}
```

```r
at_45 = at_45 %>%
  rowwise() %>%
  mutate(Home_Score = home_score_at_45(Season, Match),
         Away_Score = away_score_at_45(Season, Match),
         abs_dif = abs(Home_Score - Away_Score))
```

```r
tmp_00 = at_45 %>%
  filter(abs_dif == 0) %>%
  select(Season, Match)

HDA_00 = HDA %>%
  inner_join(tmp_00)
```

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

```r
nrow(HDA_00)
```

```
## [1] 838
```

```r
results_00 = tibble(GeoMean = apply(HDA_00[,c(99:128)], 2, EnvStats::geoMean),
                    Minuto = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
```

```r
                Modelo = factor(c(rep("A", 6),
                                  rep("B", 6),
                                  rep("C", 6),
                                  rep("D", 6),
                                  rep("E", 6)))) %>%
  pivot_wider(names_from = "Modelo", values_from = "GeoMean", names_prefix = "Modelo ") %>%
  mutate(`Modelo B` = log(`Modelo B`) - log(`Modelo A`),
         `Modelo C` = log(`Modelo C`) - log(`Modelo A`),
         `Modelo D` = log(`Modelo D`) - log(`Modelo A`),
         `Modelo E` = log(`Modelo E`) - log(`Modelo A`),
         `Modelo A` = 0) %>%
  pivot_longer(cols = starts_with("Modelo"), names_to = "Modelo", values_to = "GeoMean") %>%
  mutate(Modelo = stringr::str_replace_all(Modelo, "Modelo ", ""))

results_00 %>%
  ggplot(aes(x = Minuto, y = GeoMean, col = Modelo)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with a draw at Minuto 45") +
  ylab("Geometric mean for the results")
```


All predicted matches with a draw at Minuto 45
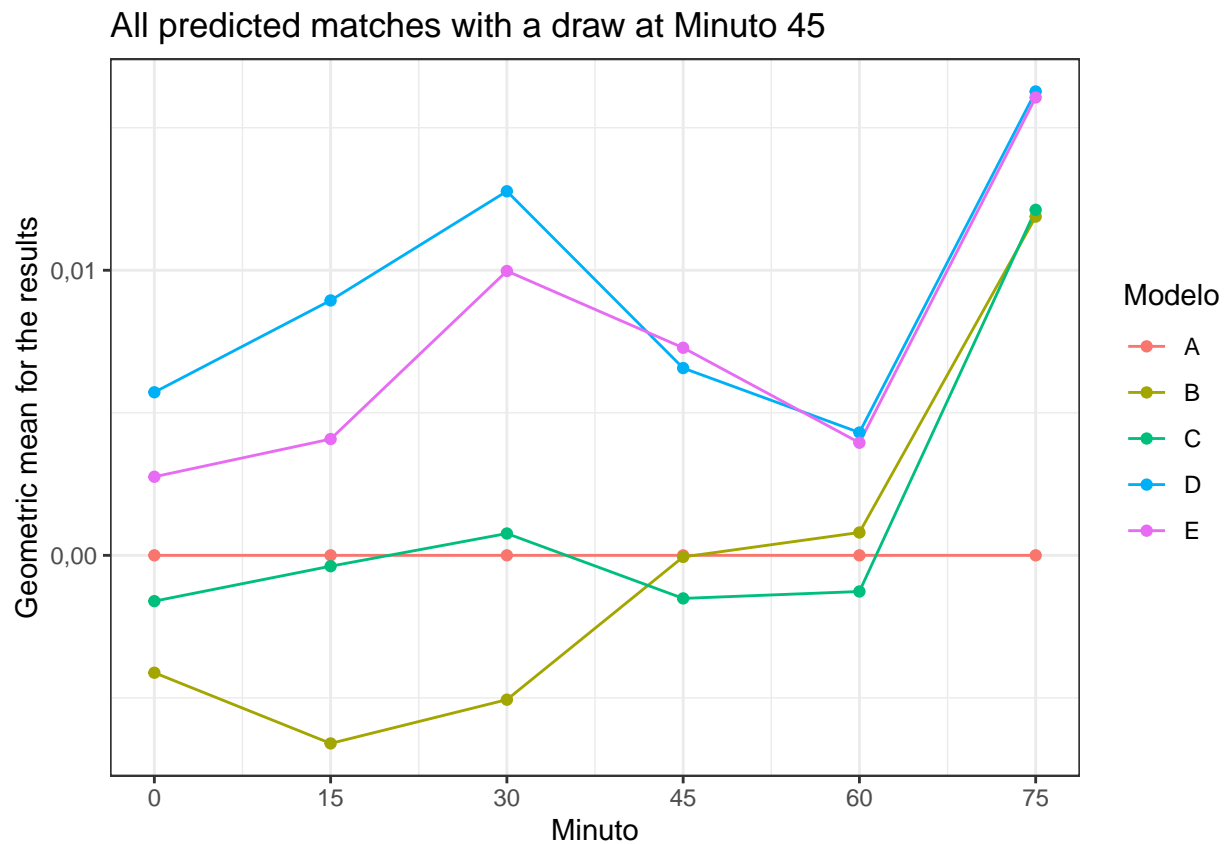
```r
results_00 %>%
  pivot_wider(id_cols = "Modelo", values_from = "GeoMean", names_from = "Minuto",
```

```
            names_prefix = "Minuto ") %>%
  kable()
```

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
A	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000
B	-0,0041180	-0,0065978	-0,0050627	-0,0000553	0,0008013	0,0118811
C	-0,0016062	-0,0003809	0,0007661	-0,0015107	-0,0012684	0,0121196
D	0,0057201	0,0089421	0,0127716	0,0065683	0,0043080	0,0162713
E	0,0027566	0,0040785	0,0099731	0,0072810	0,0039499	0,0160618

```
scores_00 = tibble(GeoMean = apply(HDA_00[,c(129:158)], 2, EnvStats::geoMean),
                   Minuto = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                   Modelo = factor(c(rep("A", 6),
                                     rep("B", 6),
                                     rep("C", 6),
                                     rep("D", 6),
                                     rep("E", 6)))) %>%
  pivot_wider(names_from = "Modelo", values_from = "GeoMean", names_prefix = "Modelo ") %>%
  mutate(`Modelo B` = log(`Modelo B`) - log(`Modelo A`),
         `Modelo C` = log(`Modelo C`) - log(`Modelo A`),
         `Modelo D` = log(`Modelo D`) - log(`Modelo A`),
         `Modelo E` = log(`Modelo E`) - log(`Modelo A`),
         `Modelo A` = 0) %>%
  pivot_longer(cols = starts_with("Modelo"), names_to = "Modelo", values_to = "GeoMean") %>%
  mutate(Modelo = stringr::str_replace_all(Modelo, "Modelo ", ""))

scores_00 %>%
  ggplot(aes(x = Minuto, y = GeoMean, col = Modelo)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with a draw at Minuto 45") +
  ylab("Geometric mean for the scores")
```

All predicted matches with a draw at Minuto 45
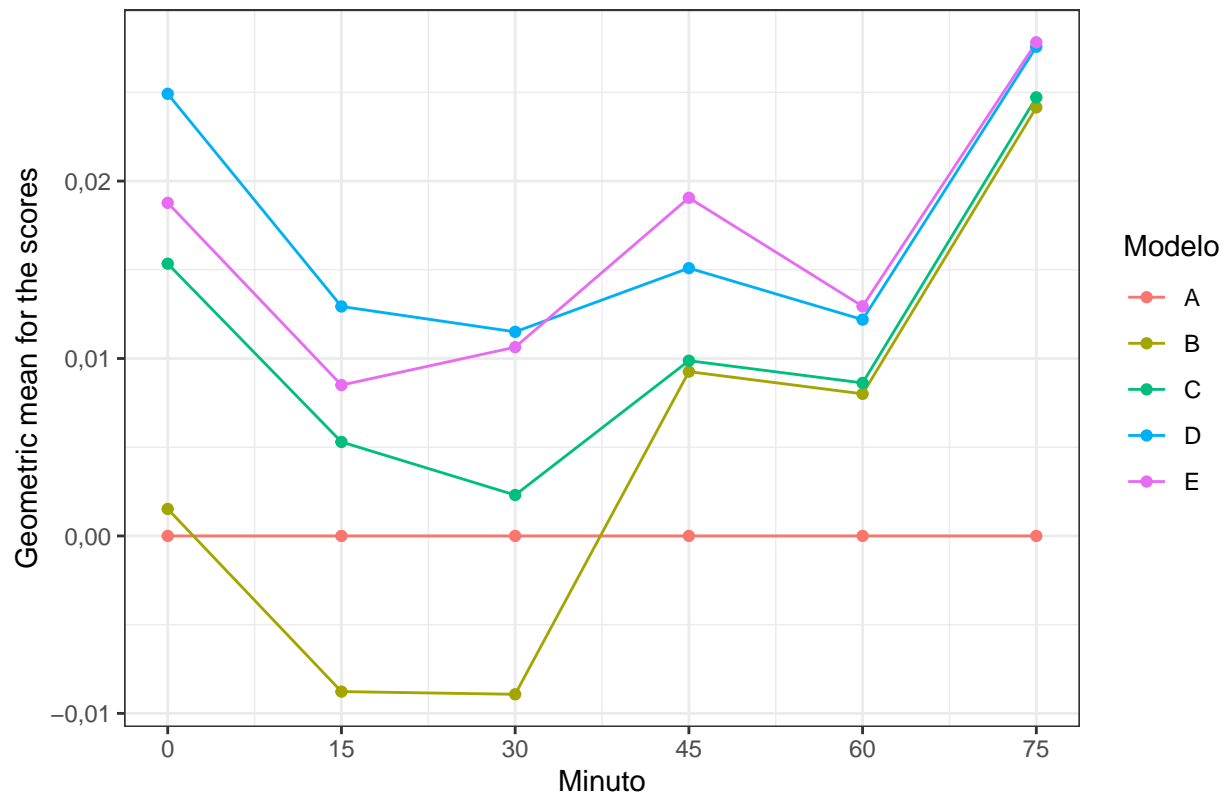

```
scores_00 %>%
  pivot_wider(id_cols = "Modelo", values_from = "GeoMean", names_from = "Minuto",
              names_prefix = "Minuto ") %>%
  kable()
```

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
A	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000
B	0,0015206	-0,0087730	-0,0089253	0,0092611	0,0080067	0,0241499
C	0,0153431	0,0052966	0,0023057	0,0098753	0,0086225	0,0247159
D	0,0249162	0,0129321	0,0115032	0,0150926	0,0121918	0,0275698
E	0,0187736	0,0085019	0,0106375	0,0190577	0,0129421	0,0278185

```
tmp_20 = at_45 %>%
  filter(abs_dif >= 2) %>%
  select(Season, Match)

HDA_20 = HDA %>%
  inner_join(tmp_20)
```

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

```
nrow(HDA_20)
```

```
## [1] 211
```

```
results_20 = tibble(GeoMean = apply(HDA_20[,c(99:128)], 2, EnvStats::geoMean),
                    Minuto = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                    Modelo = factor(c(rep("A", 6),
                                      rep("B", 6),
                                      rep("C", 6),
                                      rep("D", 6),
                                      rep("E", 6)))) %>%
  pivot_wider(names_from = "Modelo", values_from = "GeoMean", names_prefix = "Modelo ") %>%
  mutate(`Modelo B` = log(`Modelo B`) - log(`Modelo A`),
         `Modelo C` = log(`Modelo C`) - log(`Modelo A`),
         `Modelo D` = log(`Modelo D`) - log(`Modelo A`),
         `Modelo E` = log(`Modelo E`) - log(`Modelo A`),
         `Modelo A` = 0) %>%
  pivot_longer(cols = starts_with("Modelo"), names_to = "Modelo", values_to = "GeoMean") %>%
  mutate(Modelo = stringr::str_replace_all(Modelo, "Modelo ", ""))

results_20 %>%
  ggplot(aes(x = Minuto, y = GeoMean, col = Modelo)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with a 2+ goal lead at Minuto 45") +
  ylab("Geometric mean for the results")
```

All predicted matches with a 2+ goal lead at Minuto 45
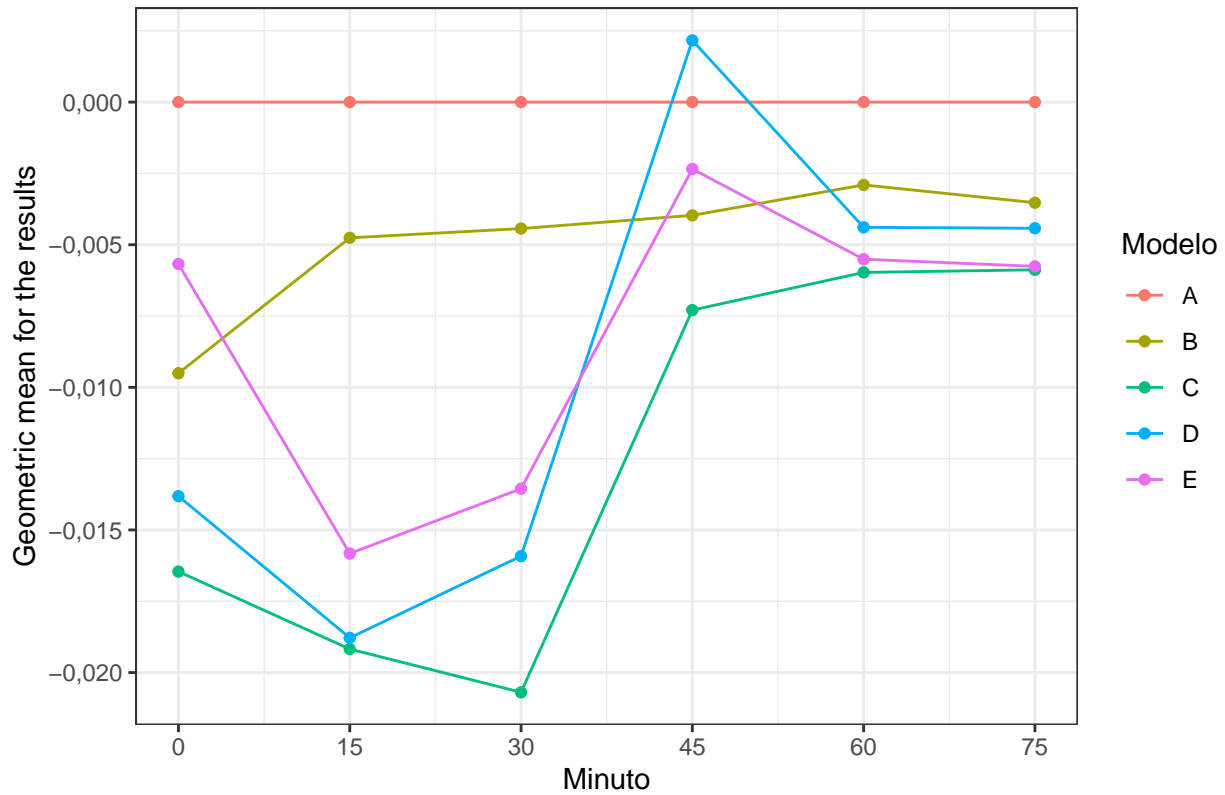

```
results_20 %>%
  pivot_wider(id_cols = "Modelo", values_from = "GeoMean", names_from = "Minuto",
              names_prefix = "Minuto ") %>%
  kable()
```

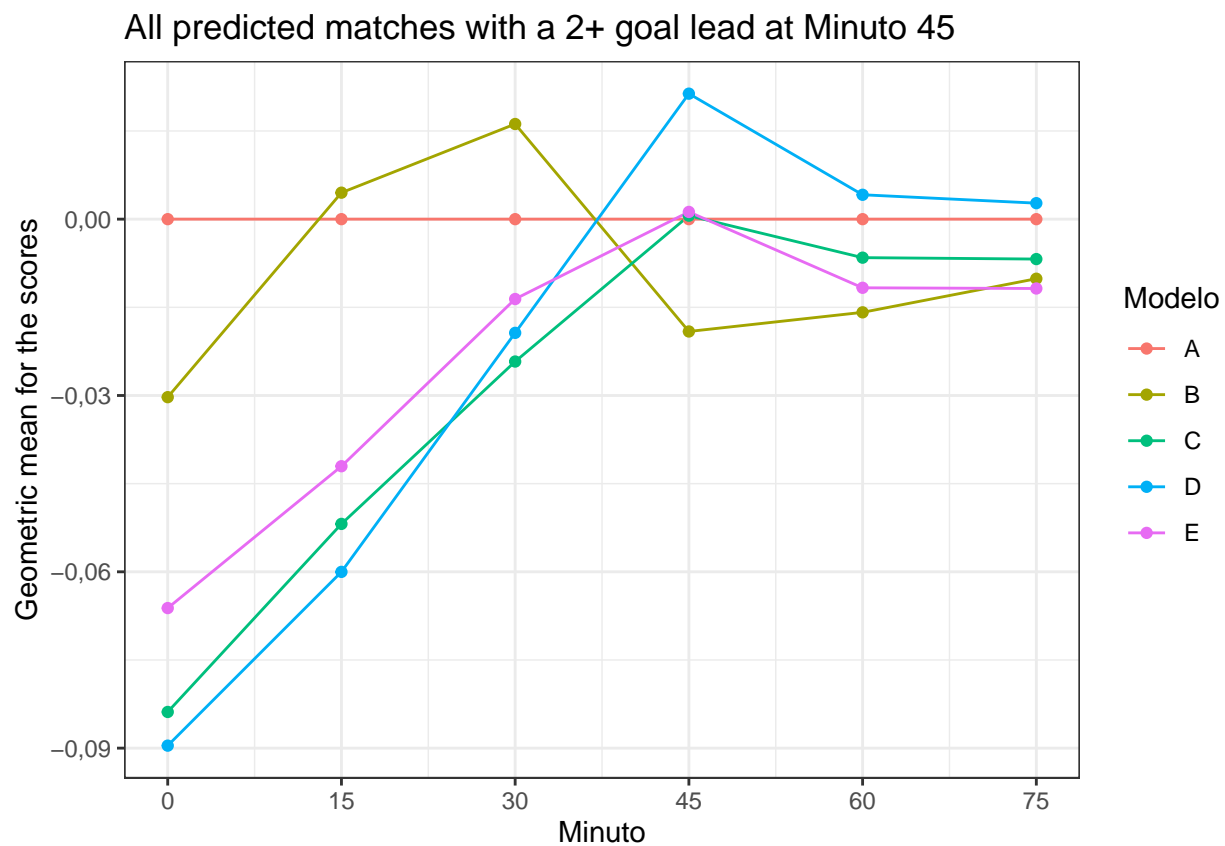| Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
A	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000
B	-0,0095059	-0,0047586	-0,0044337	-0,0039713	-0,0029034	-0,0035282
C	-0,0164595	-0,0191762	-0,0206926	-0,0072930	-0,0059716	-0,0058801
D	-0,0138164	-0,0187832	-0,0159184	0,0021653	-0,0043905	-0,0044233
E	-0,0056755	-0,0158261	-0,0135563	-0,0023447	-0,0055092	-0,0057598

```
scores_20 = tibble(GeoMean = apply(HDA_20[,c(129:158)], 2, EnvStats::geoMean),
                   Minuto = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                   Modelo = factor(c(rep("A", 6),
                                     rep("B", 6),
                                     rep("C", 6),
                                     rep("D", 6),
                                     rep("E", 6)))) %>%
  pivot_wider(names_from = "Modelo", values_from = "GeoMean", names_prefix = "Modelo ") %>%
  mutate(`Modelo B` = log(`Modelo B`) - log(`Modelo A`),
         `Modelo C` = log(`Modelo C`) - log(`Modelo A`),
         `Modelo D` = log(`Modelo D`) - log(`Modelo A`),
```

```
        `Modelo E` = log(`Modelo E`) - log(`Modelo A`),
        `Modelo A` = 0) %>%
  pivot_longer(cols = starts_with("Modelo"), names_to = "Modelo", values_to = "GeoMean") %>%
  mutate(Modelo = stringr::str_replace_all(Modelo, "Modelo ", ""))

scores_20 %>%
  ggplot(aes(x = Minuto, y = GeoMean, col = Modelo)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with a 2+ goal lead at Minuto 45") +
  ylab("Geometric mean for the scores")
```



```
scores_20 %>%
  pivot_wider(id_cols = "Modelo", values_from = "GeoMean", names_from = "Minuto",
              names_prefix = "Minuto ") %>%
  kable()
```

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
A	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000
B	-0,0302863	0,0045028	0,0161864	-0,0191086	-0,0158537	-0,0101350
C	-0,0838447	-0,0518512	-0,0242255	0,0005933	-0,0065506	-0,0067907
D	-0,0895750	-0,0600116	-0,0193468	0,0213497	0,0041478	0,0027224

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
E	-0,0661697	-0,0420375	-0,0135868	0,0012302	-0,0116800	-0,0117986

```
tmp_10 = at_45 %>%
  filter(Home_Score == 1, Away_Score == 0) %>%
  select(Season, Match)

HDA_10 = HDA %>%
  inner_join(tmp_10)
```

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

```
nrow(HDA_10)
```

```
## [1] 435
```

```
results_10 = tibble(GeoMean = apply(HDA_10[,c(99:128)], 2, EnvStats::geoMean),
                    Minuto = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                    Modelo = factor(c(rep("A", 6),
                                      rep("B", 6),
                                      rep("C", 6),
                                      rep("D", 6),
                                      rep("E", 6)))) %>%
  pivot_wider(names_from = "Modelo", values_from = "GeoMean", names_prefix = "Modelo ") %>%
  mutate(`Modelo B` = log(`Modelo B`) - log(`Modelo A`),
         `Modelo C` = log(`Modelo C`) - log(`Modelo A`),
         `Modelo D` = log(`Modelo D`) - log(`Modelo A`),
         `Modelo E` = log(`Modelo E`) - log(`Modelo A`),
         `Modelo A` = 0) %>%
  pivot_longer(cols = starts_with("Modelo"), names_to = "Modelo", values_to = "GeoMean") %>%
  mutate(Modelo = stringr::str_replace_all(Modelo, "Modelo ", ""))

results_10 %>%
  ggplot(aes(x = Minuto, y = GeoMean, col = Modelo)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with score 1-0 at Minuto 45") +
  ylab("Geometric mean for the results")
```

All predicted matches with score 1−0 at Minuto 45
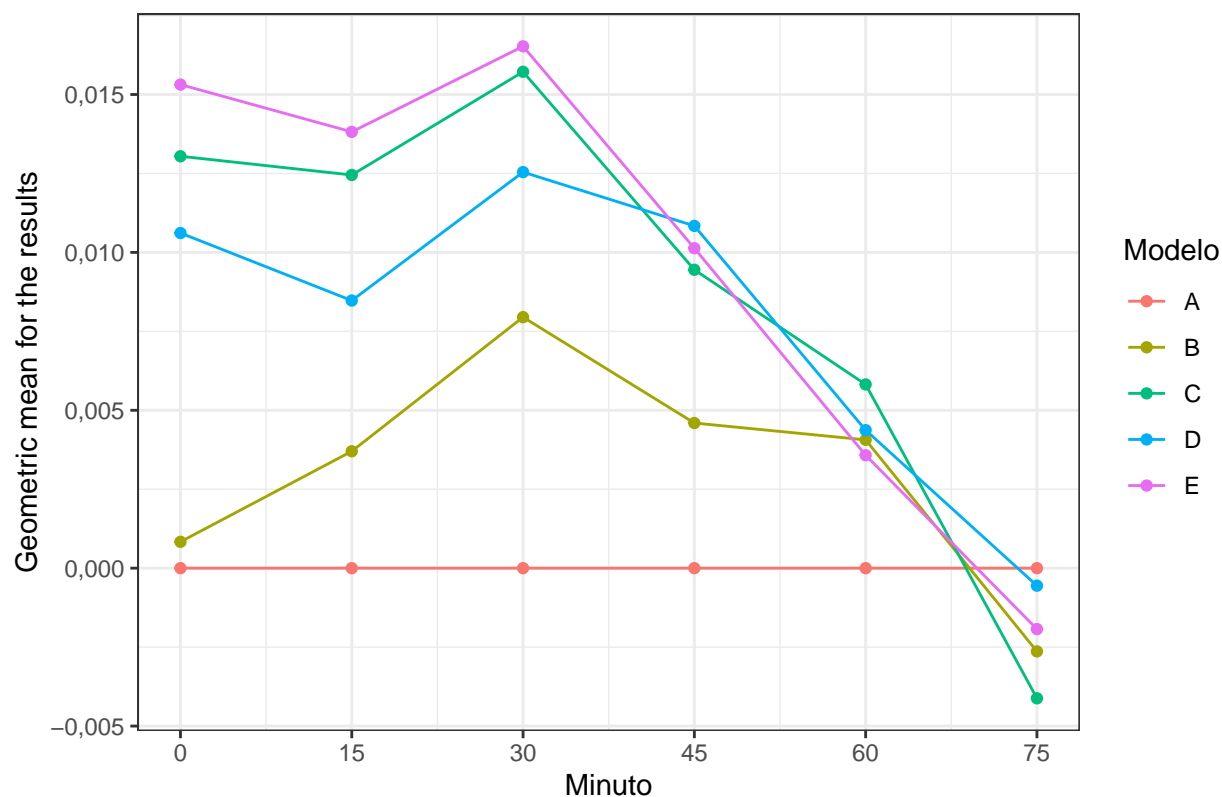

```
results_10 %>%
  pivot_wider(id_cols = "Modelo", values_from = "GeoMean", names_from = "Minuto",
              names_prefix = "Minuto ") %>%
  kable()
```

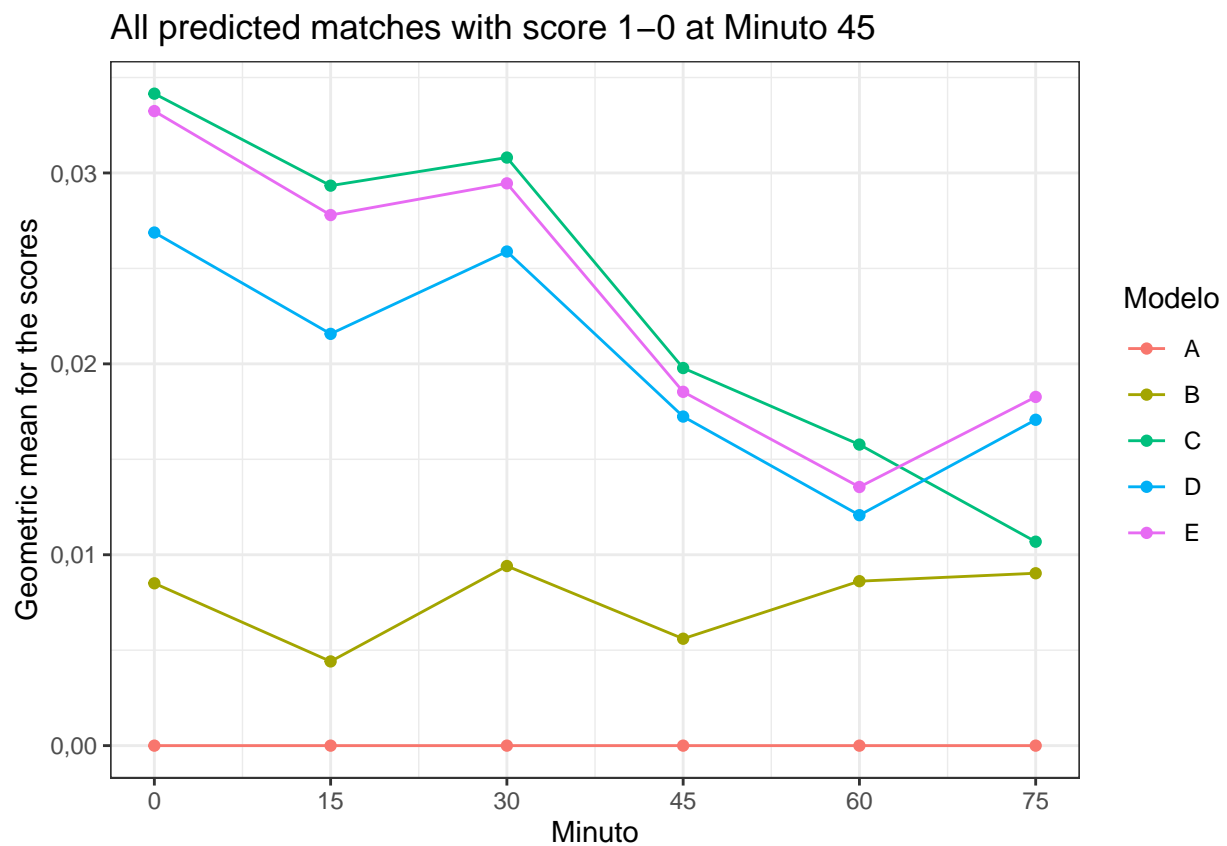| Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
A	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000
B	0,0008345	0,0037022	0,0079499	0,0045958	0,0040635	-0,0026358
C	0,0130443	0,0124519	0,0157191	0,0094499	0,0058181	-0,0041212
D	0,0106138	0,0084752	0,0125425	0,0108392	0,0043706	-0,0005552
E	0,0153144	0,0138165	0,0165257	0,0101325	0,0035766	-0,0019285

```
scores_10 = tibble(GeoMean = apply(HDA_10[,c(129:158)], 2, EnvStats::geoMean),
                   Minuto = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                   Modelo = factor(c(rep("A", 6),
                                     rep("B", 6),
                                     rep("C", 6),
                                     rep("D", 6),
                                     rep("E", 6)))) %>%
  pivot_wider(names_from = "Modelo", values_from = "GeoMean", names_prefix = "Modelo ") %>%
  mutate(`Modelo B` = log(`Modelo B`) - log(`Modelo A`),
         `Modelo C` = log(`Modelo C`) - log(`Modelo A`),
         `Modelo D` = log(`Modelo D`) - log(`Modelo A`),
```

```
        `Modelo E` = log(`Modelo E`) - log(`Modelo A`),
        `Modelo A` = 0) %>%
  pivot_longer(cols = starts_with("Modelo"), names_to = "Modelo", values_to = "GeoMean") %>%
  mutate(Modelo = stringr::str_replace_all(Modelo, "Modelo ", ""))

scores_10 %>%
  ggplot(aes(x = Minuto, y = GeoMean, col = Modelo)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with score 1-0 at Minuto 45") +
  ylab("Geometric mean for the scores")
```



```
scores_10 %>%
  pivot_wider(id_cols = "Modelo", values_from = "GeoMean", names_from = "Minuto",
              names_prefix = "Minuto ") %>%
  kable()
```

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
A	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000
B	0,0085043	0,0044119	0,0094102	0,0056006	0,0086154	0,0090305
C	0,0341541	0,0293356	0,0308079	0,0197809	0,0157713	0,0106837
D	0,0268790	0,0215709	0,0258845	0,0172387	0,0120726	0,0170711

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
E	0,0332421	0,0277955	0,0294563	0,0185350	0,0135493	0,0182678

```r
tmp_01 = at_45 %>%
  filter(Home_Score == 0, Away_Score == 1) %>%
  select(Season, Match)

HDA_01 = HDA %>%
  inner_join(tmp_01)
```

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

```r
nrow(HDA_01)
```

```
## [1] 275
```

```r
results_01 = tibble(GeoMean = apply(HDA_01[,c(99:128)], 2, EnvStats::geoMean),
                    Minuto = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                    Modelo = factor(c(rep("A", 6),
                                      rep("B", 6),
                                      rep("C", 6),
                                      rep("D", 6),
                                      rep("E", 6)))) %>%
  pivot_wider(names_from = "Modelo", values_from = "GeoMean", names_prefix = "Modelo ") %>%
  mutate(`Modelo B` = log(`Modelo B`) - log(`Modelo A`),
         `Modelo C` = log(`Modelo C`) - log(`Modelo A`),
         `Modelo D` = log(`Modelo D`) - log(`Modelo A`),
         `Modelo E` = log(`Modelo E`) - log(`Modelo A`),
         `Modelo A` = 0) %>%
  pivot_longer(cols = starts_with("Modelo"), names_to = "Modelo", values_to = "GeoMean") %>%
  mutate(Modelo = stringr::str_replace_all(Modelo, "Modelo ", ""))

results_01 %>%
  ggplot(aes(x = Minuto, y = GeoMean, col = Modelo)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with score 0-1 at Minuto 45") +
  ylab("Geometric mean for the results")
```

All predicted matches with score 0–1 at Minuto 45
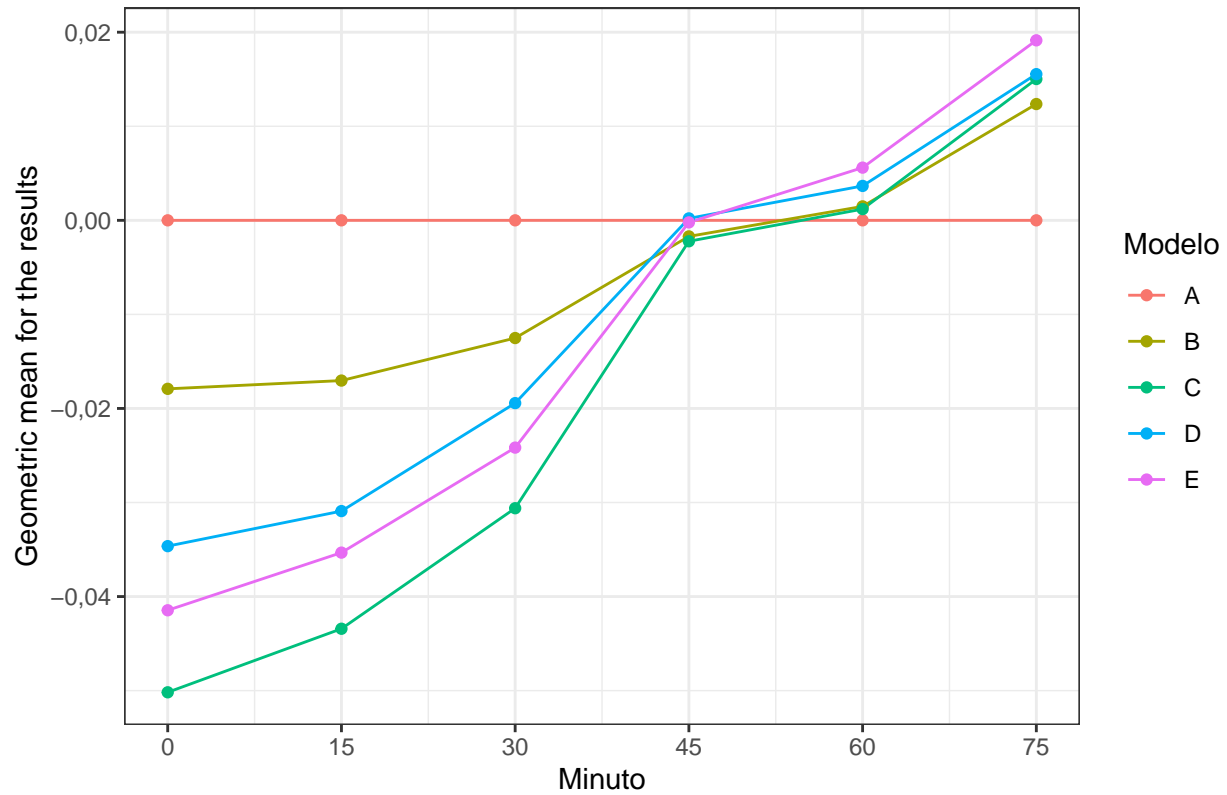

```r
results_01 %>%
  pivot_wider(id_cols = "Modelo", values_from = "GeoMean", names_from = "Minuto",
              names_prefix = "Minuto ") %>%
  kable()
```

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
A	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000
B	-0,0179159	-0,0170361	-0,0125142	-0,0017080	0,0014821	0,0123630
C	-0,0501786	-0,0434216	-0,0306085	-0,0022241	0,0011931	0,0150233
D	-0,0346381	-0,0309120	-0,0194354	0,0002052	0,0036562	0,0155435
E	-0,0414568	-0,0353285	-0,0241641	-0,0002188	0,0056020	0,0191365

```r
scores_01 = tibble(GeoMean = apply(HDA_01[,c(129:158)], 2, EnvStats::geoMean),
                   Minuto = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                   Modelo = factor(c(rep("A", 6),
                                     rep("B", 6),
                                     rep("C", 6),
                                     rep("D", 6),
                                     rep("E", 6)))) %>%
  pivot_wider(names_from = "Modelo", values_from = "GeoMean", names_prefix = "Modelo ") %>%
  mutate(`Modelo B` = log(`Modelo B`) - log(`Modelo A`),
         `Modelo C` = log(`Modelo C`) - log(`Modelo A`),
         `Modelo D` = log(`Modelo D`) - log(`Modelo A`),
```

```
        `Modelo E` = log(`Modelo E`) - log(`Modelo A`),
        `Modelo A` = 0) %>%
  pivot_longer(cols = starts_with("Modelo"), names_to = "Modelo", values_to = "GeoMean") %>%
  mutate(Modelo = stringr::str_replace_all(Modelo, "Modelo ", ""))

scores_01 %>%
  ggplot(aes(x = Minuto, y = GeoMean, col = Modelo)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with score 0-1 at Minuto 45") +
  ylab("Geometric mean for the scores")
```
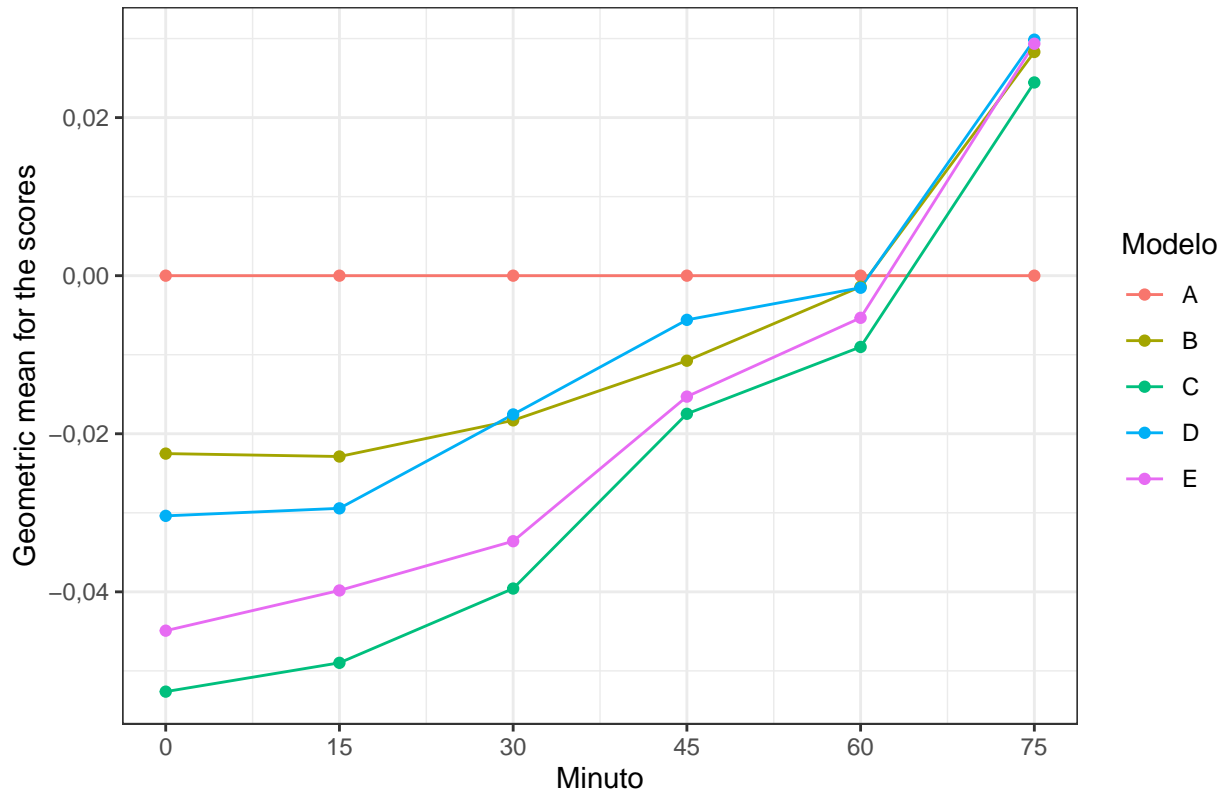


```
scores_01 %>%
  pivot_wider(id_cols = "Modelo", values_from = "GeoMean", names_from = "Minuto",
              names_prefix = "Minuto ") %>%
  kable()
```

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
A	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000
B	-0,0225112	-0,0228786	-0,0183044	-0,0107573	-0,0013956	0,0282999
C	-0,0526267	-0,0489820	-0,0395837	-0,0174752	-0,0090168	0,0244431
D	-0,0303868	-0,0294374	-0,0175572	-0,0055894	-0,0015256	0,0298584

Modelo	Minuto 0	Minuto 15	Minuto 30	Minuto 45	Minuto 60	Minuto 75
E	-0,0449157	-0,0398251	-0,0335938	-0,0153024	-0,0053331	0,0293473