

Geometric mean for the scores

```
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)

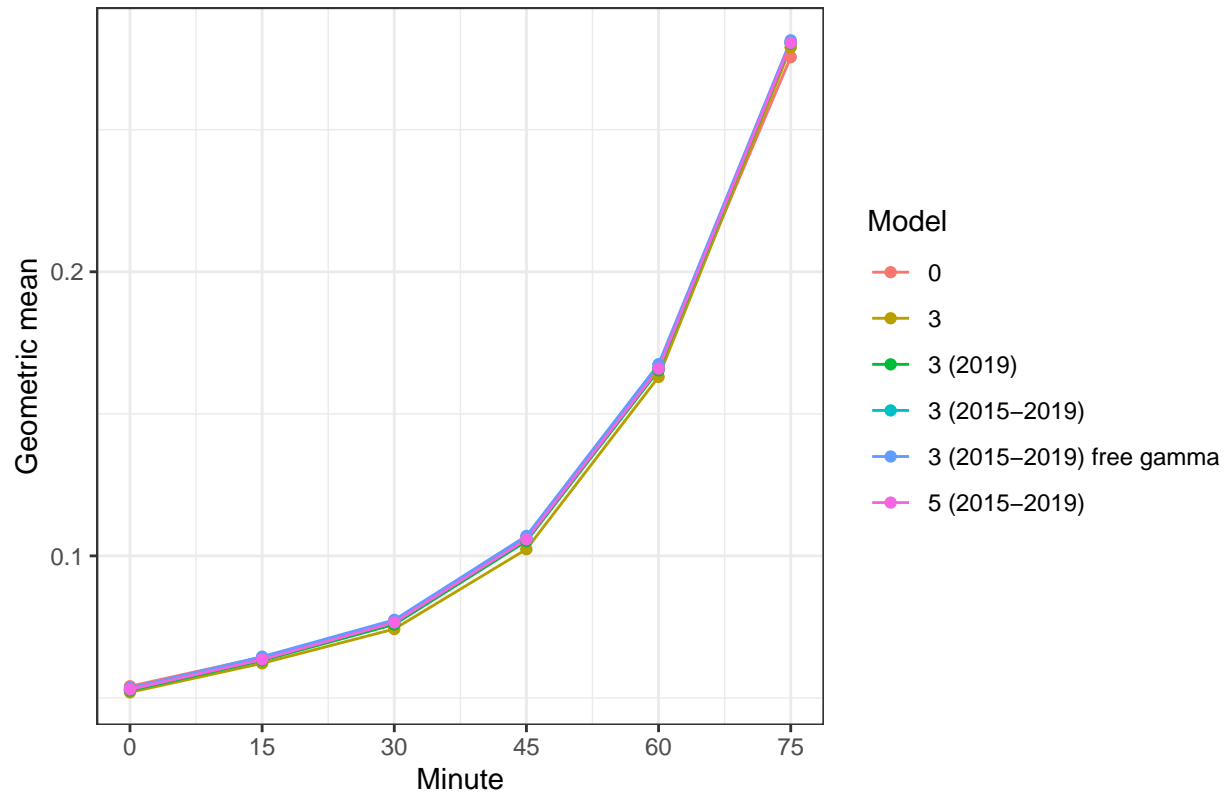
## [1] 333

HDA2[which(HDA2 == 0, arr.ind = TRUE)] = 10^-5 # pra não ter NA (tinha em dois casos)

all = tibble(RPS = apply(HDA2[,c(255:290)], 2, EnvStats::geoMean),
             Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 6)),
             Model = factor(c(rep("0", 6),
                               rep("3", 6),
                               rep("3 (2019)", 6),
                               rep("3 (2015-2019)", 6),
                               rep("3 (2015-2019) free gamma", 6),
                               rep("5 (2015-2019)", 6)),
                             levels = c("0",
                                         "3",
                                         "3 (2019)",
                                         "3 (2015-2019)",
                                         "3 (2015-2019) free gamma",
                                         "5 (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	0.0541099	0.0643832	0.0769402	0.1059196	0.1659344	0.2755738
3	0.0519795	0.0621783	0.0742501	0.1023422	0.1629886	0.2789613
3 (2019)	0.0526334	0.0631413	0.0758853	0.1051659	0.1653138	0.2803020
3 (2015-2019)	0.0529738	0.0636925	0.0766191	0.1059654	0.1663335	0.2809613
3 (2015-2019) free gamma	0.0535071	0.0645129	0.0774704	0.1070388	0.1674743	0.2814751
5 (2015-2019)	0.0530204	0.0635160	0.0765648	0.1058098	0.1659045	0.2805620

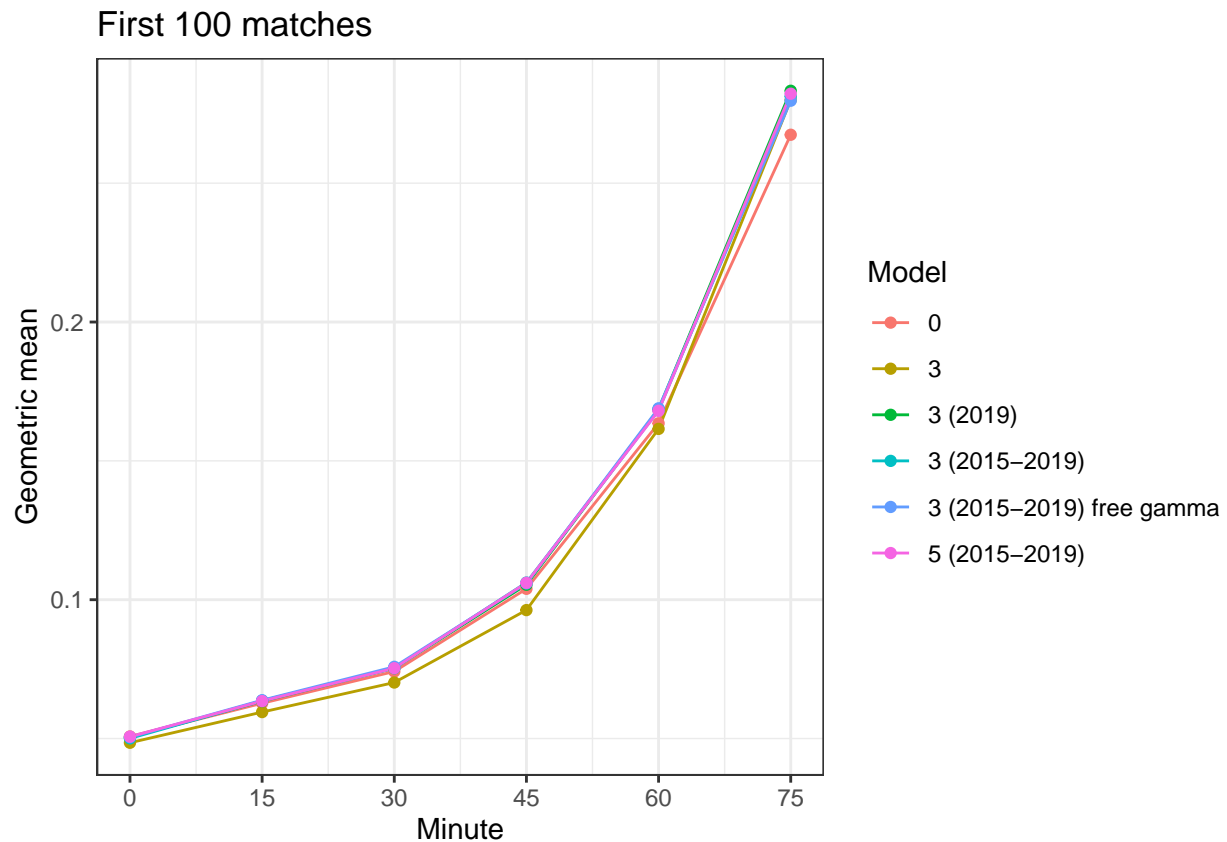
```
first_100 = tibble(RPS = apply(HDA2[c(1:100), c(255:290)], 2, EnvStats::geoMean),
  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 6)),
  Model = factor(c(rep("0", 6),
    rep("3", 6),
    rep("3 (2019)", 6),
    rep("3 (2015-2019)", 6),
    rep("3 (2015-2019) free gamma", 6),
    rep("5 (2015-2019)", 6)),
  levels = c("0",
    "3",
```

```

"3 (2019)",
"3 (2015-2019)",
"3 (2015-2019) free gamma",
"5 (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	0.0507481	0.0627055	0.0741431	0.1039253	0.1635207	0.2674356
3	0.0484814	0.0595347	0.0701730	0.0962585	0.1615151	0.2799821
3 (2019)	0.0504495	0.0634178	0.0753294	0.1053927	0.1686049	0.2832706
3 (2015-2019)	0.0500492	0.0636221	0.0753647	0.1061667	0.1681872	0.2812942
3 (2015-2019) free gamma	0.0504649	0.0638020	0.0758302	0.1058801	0.1689199	0.2796967

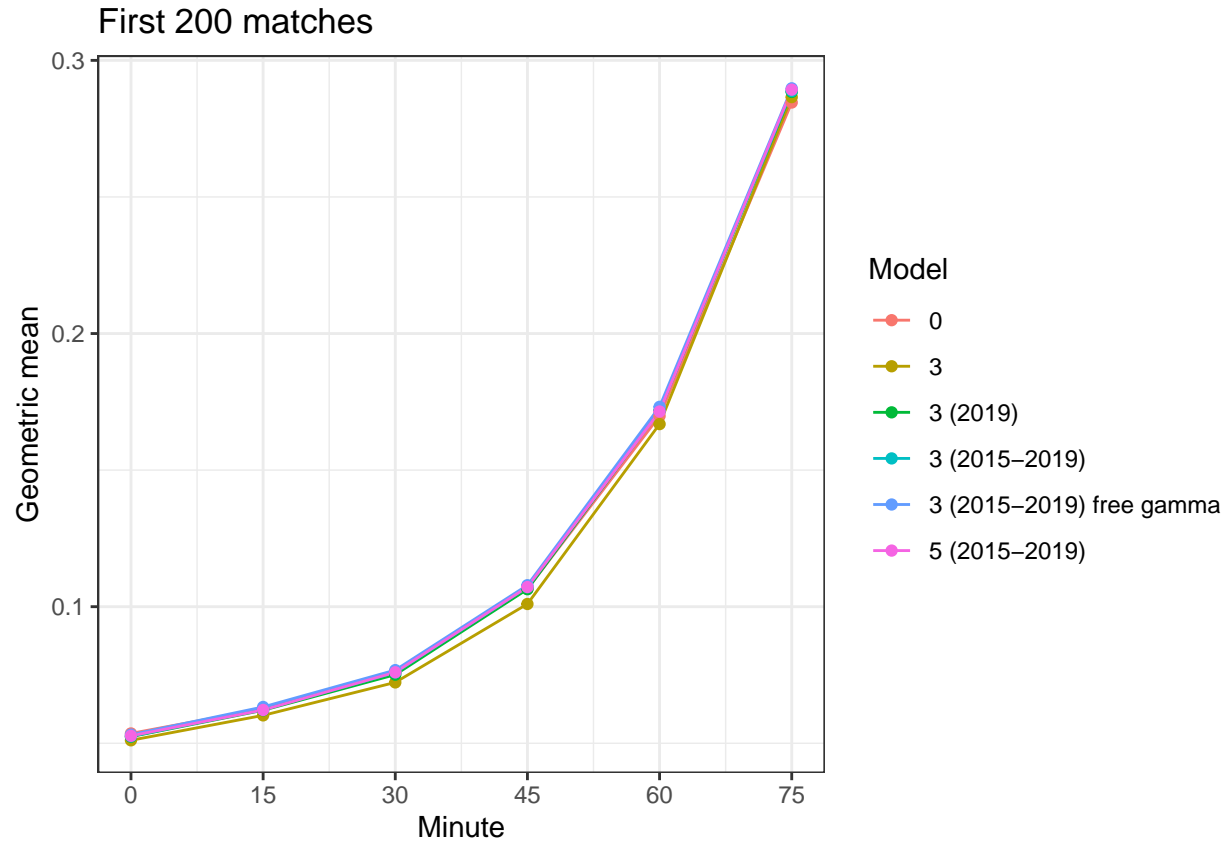
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
5 (2015-2019)	0.0506829	0.0634764	0.0753250	0.1060804	0.1680215	0.2822210

```

first_200 = tibble(RPS = apply(HDA2[c(1:200), c(255:290)], 2, EnvStats::geoMean),
  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 6)),
  Model = factor(c(rep("0", 6),
    rep("3", 6),
    rep("3 (2019)", 6),
    rep("3 (2015-2019)", 6),
    rep("3 (2015-2019) free gamma" , 6),
    rep("5 (2015-2019)", 6)),
    levels = c("0",
      "3",
      "3 (2019)",
      "3 (2015-2019)",
      "3 (2015-2019) free gamma",
      "5 (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	0.0536038	0.0626504	0.0762190	0.1065319	0.1698249	0.2845741
3	0.0510903	0.0602133	0.0722861	0.1009975	0.1668495	0.2865561
3 (2019)	0.0524347	0.0620882	0.0751212	0.1064063	0.1717093	0.2884439
3 (2015-2019)	0.0525564	0.0623981	0.0759135	0.1071815	0.1718088	0.2887681
3 (2015-2019) free gamma	0.0531553	0.0632976	0.0767884	0.1078631	0.1731616	0.2897618
5 (2015-2019)	0.0527352	0.0621904	0.0759870	0.1071908	0.1714583	0.2892770

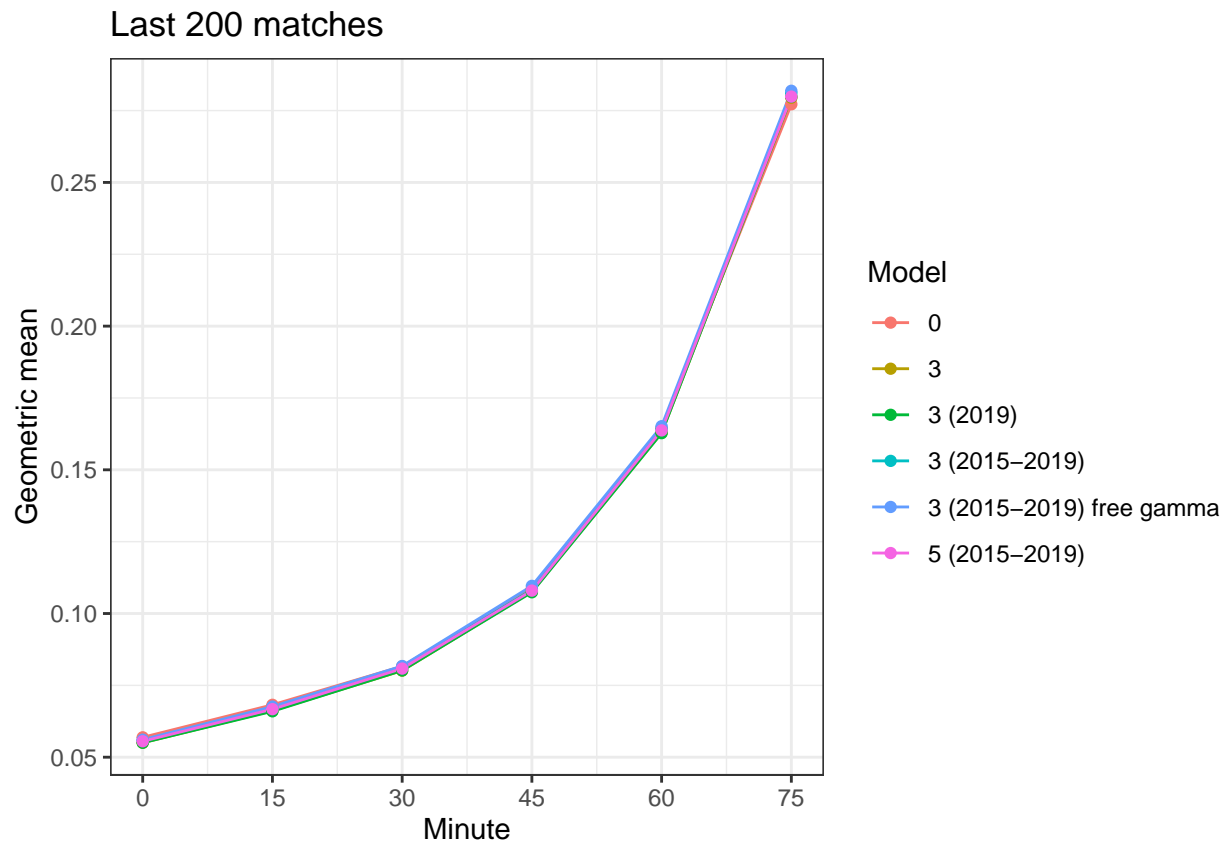
```
last_200 = tibble(RPS = apply(HDA2[c(134:333), c(255:290)], 2, EnvStats::geoMean),
                  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 6)),
                  Model = factor(c(rep("0", 6),
                                   rep("3", 6),
                                   rep("3 (2019)", 6),
                                   rep("3 (2015-2019)", 6),
                                   rep("3 (2015-2019) free gamma", 6),
                                   rep("5 (2015-2019)", 6)),
                                levels = c("0",
                                             "3",
```

```

"3 (2019)",
"3 (2015-2019)",
"3 (2015-2019) free gamma",
"5 (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()

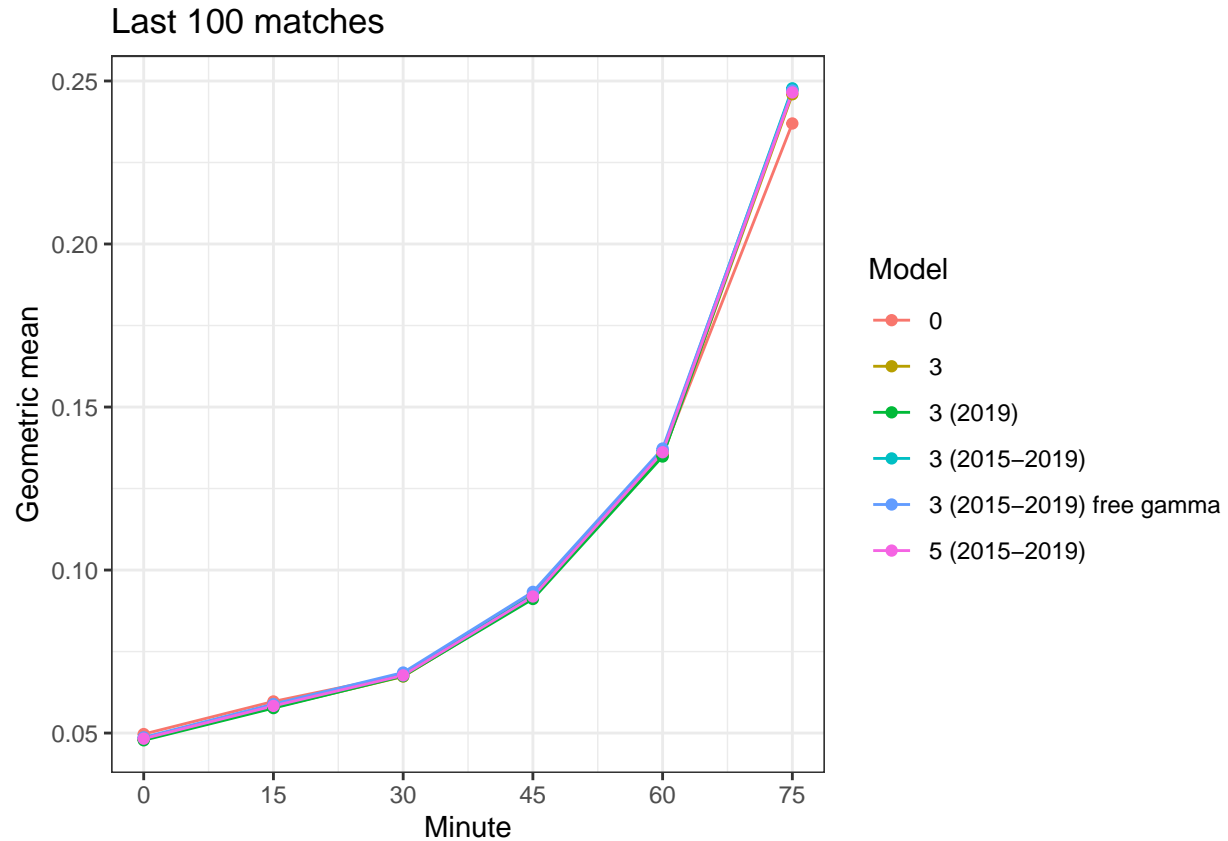
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.0569159	0.0682275	0.0816350	0.1087268	0.1645586	0.2771887
3	0.0550259	0.0661647	0.0801317	0.1076446	0.1628846	0.2794829
3 (2019)	0.0550336	0.0659689	0.0801899	0.1073916	0.1627663	0.2798505
3 (2015-2019)	0.0557518	0.0668327	0.0810401	0.1082346	0.1642160	0.2811035
3 (2015-2019) free gamma	0.0561026	0.0676082	0.0817225	0.1096139	0.1651796	0.2818889

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
5 (2015-2019)	0.0555431	0.0667161	0.0808423	0.1079620	0.1637153	0.2798695

```
last_100 = tibble(RPS = apply(HDA2[c(234:333), c(255:290)], 2, EnvStats::geoMean),
  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 6)),
  Model = factor(c(rep("0", 6),
    rep("3", 6),
    rep("3 (2019)", 6),
    rep("3 (2015-2019)", 6),
    rep("3 (2015-2019) free gamma", 6),
    rep("5 (2015-2019)", 6)),
    levels = c("0",
      "3",
      "3 (2019)",
      "3 (2015-2019)",
      "3 (2015-2019) free gamma",
      "5 (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	0.0496923	0.0596935	0.0679192	0.0925919	0.1356696	0.2369678
3	0.0480563	0.0578564	0.0673386	0.0915845	0.1350495	0.2459200
3 (2019)	0.0477805	0.0576574	0.0674782	0.0911226	0.1348228	0.2472908
3 (2015-2019)	0.0484749	0.0585068	0.0680481	0.0922018	0.1366273	0.2477461
3 (2015-2019) free gamma	0.0486800	0.0588836	0.0685656	0.0933028	0.1372633	0.2470800
5 (2015-2019)	0.0482805	0.0583274	0.0676870	0.0918921	0.1361283	0.2464883

```
matches = reds %>%
  filter(Season == 2020, Half == 1) %>%
  .$Match
length(matches)
```

```
## [1] 23
```

```
HDA2_reds = HDA2 %>%
  filter(Match %in% matches)
```

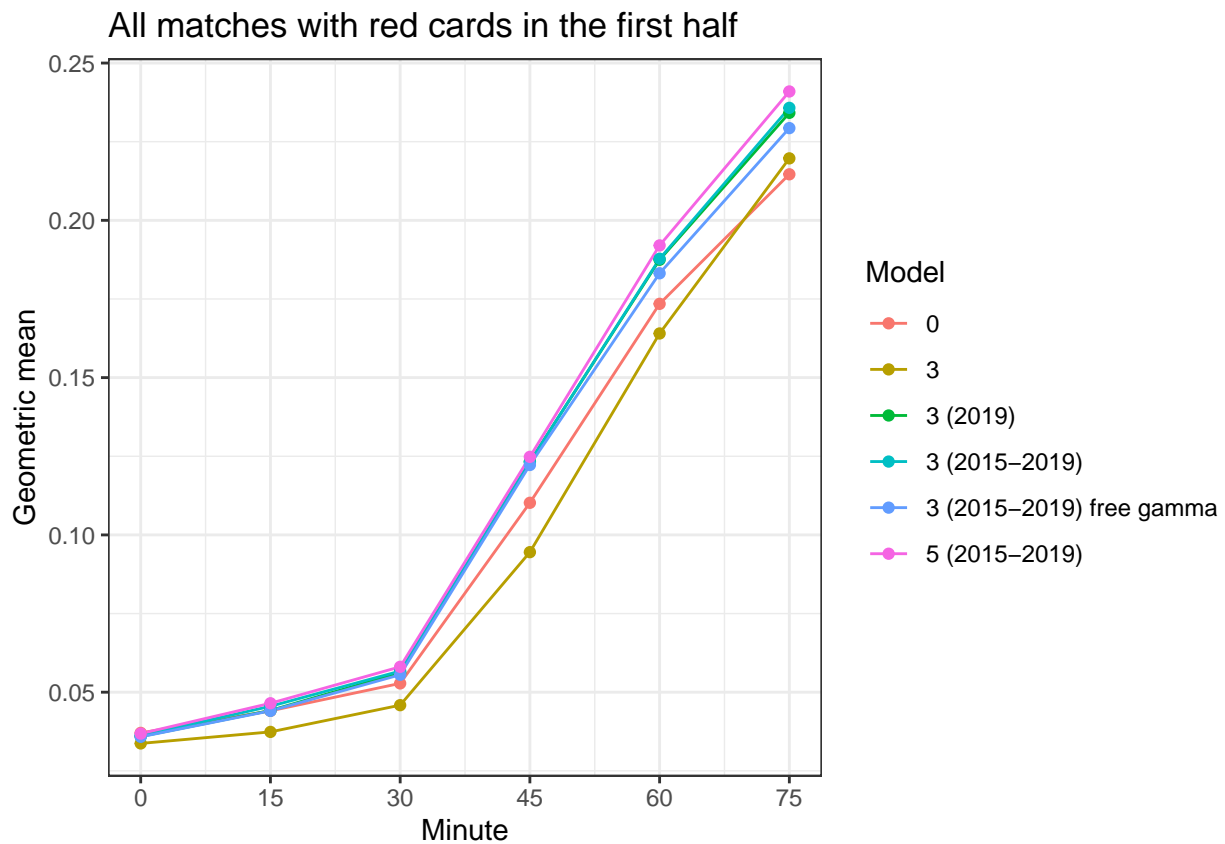


```

all_recs = tibble(RPS = apply(HDA2_recs[,c(255:290)], 2, EnvStats::geoMean),
  Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 6)),
  Model = factor(c(rep("0", 6),
    rep("3", 6),
    rep("3 (2019)", 6),
    rep("3 (2015-2019)", 6),
    rep("3 (2015-2019) free gamma", 6),
    rep("5 (2015-2019)", 6)),
  levels = c("0",
    "3",
    "3 (2019)",
    "3 (2015-2019)",
    "3 (2015-2019) free gamma",
    "5 (2015-2019)")))

all_recs %>%
  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_recs %>%
  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	0.0370633	0.0440737	0.0528048	0.1102081	0.1734562	0.2146396
3	0.0337142	0.0373789	0.0458767	0.0945157	0.1640380	0.2197000
3 (2019)	0.0359246	0.0441976	0.0561886	0.1232955	0.1874730	0.2341746
3 (2015-2019)	0.0362642	0.0455170	0.0567120	0.1231845	0.1877847	0.2357492
3 (2015-2019) free gamma	0.0358231	0.0440732	0.0555447	0.1222345	0.1831662	0.2293189
5 (2015-2019)	0.0368294	0.0464866	0.0580713	0.1248152	0.1920285	0.2409630