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] 333
```

```r
all = tibble(RPS = apply(HDA[,c(189:224)], 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
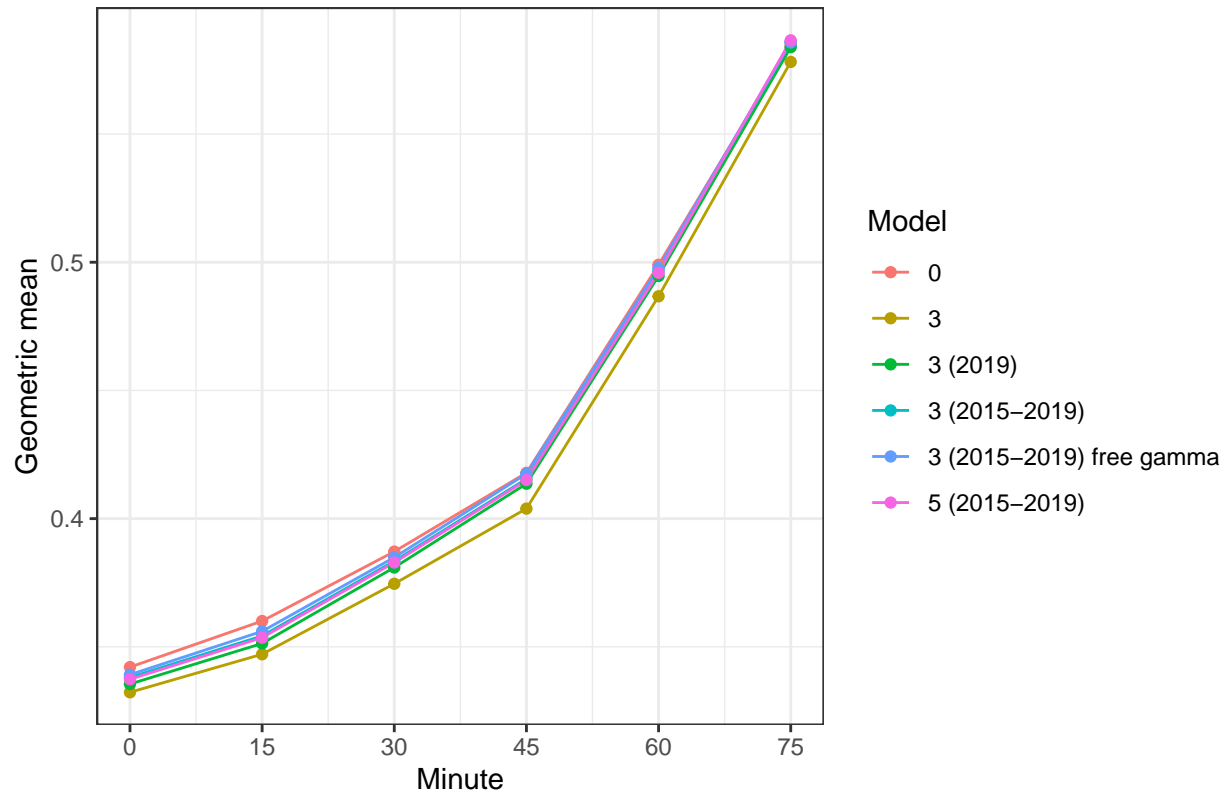

```
all %>%
  pivot_wider(id_cols = "Model", values_from = "RPS", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

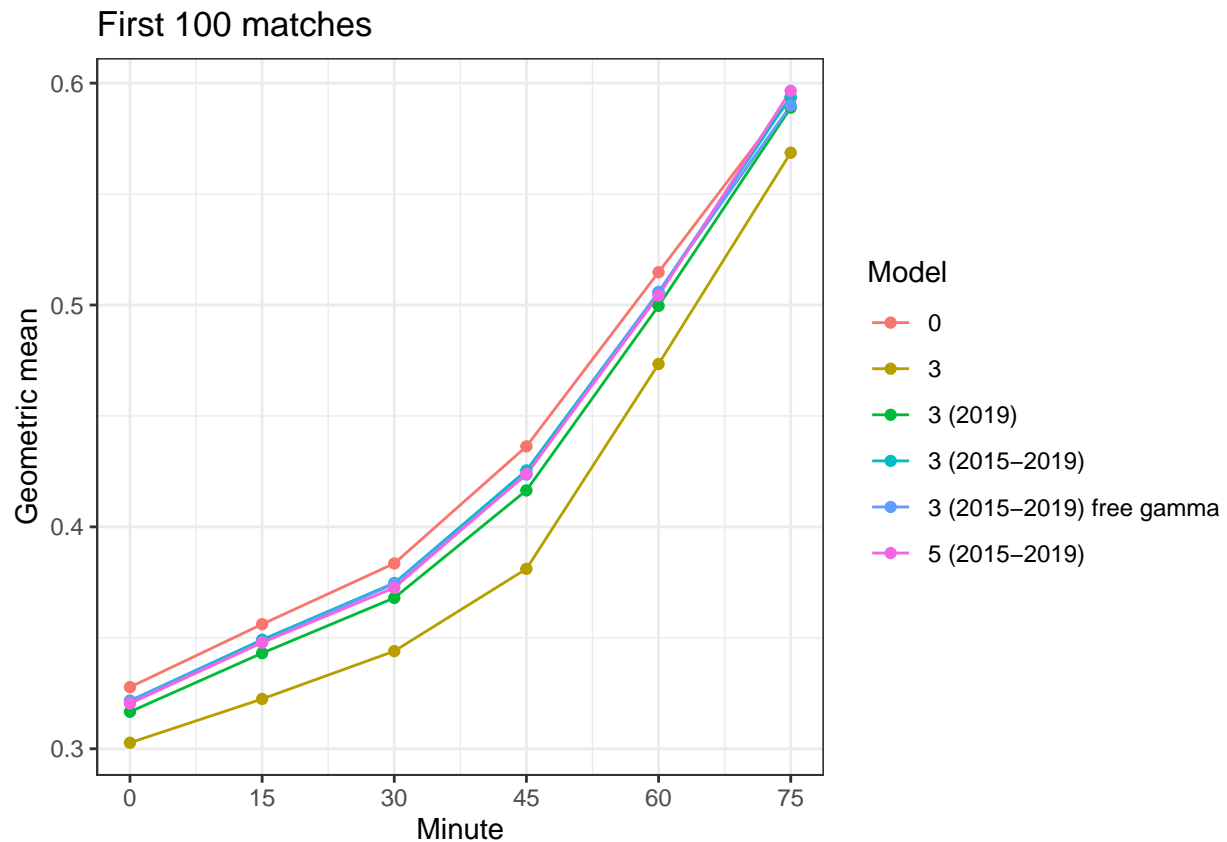| Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.3420904	0.3600642	0.3871388	0.4178542	0.4990023	0.5846716
3	0.3322657	0.3471093	0.3745867	0.4039036	0.4867176	0.5781324
3 (2019)	0.3354902	0.3513164	0.3809432	0.4136202	0.4946579	0.5838888
3 (2015-2019)	0.3380622	0.3542434	0.3834324	0.4154942	0.4966345	0.5857869
3 (2015-2019) free gamma	0.3391432	0.3560490	0.3848852	0.4174927	0.4977724	0.5857710
5 (2015-2019)	0.3373542	0.3535897	0.3828822	0.4150477	0.4958691	0.5865400

```
first_100 = tibble(RPS = apply(HDA[c(1:100), c(189:224)], 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	0.3277782	0.3561339	0.3835136	0.4362936	0.5147791	0.5932079
3	0.3027024	0.3224332	0.3439896	0.3810850	0.4734047	0.5686402
3 (2019)	0.3166519	0.3430867	0.3679943	0.4164512	0.4996328	0.5889556
3 (2015-2019)	0.3216680	0.3491274	0.3746992	0.4254306	0.5057311	0.5938677
3 (2015-2019) free gamma	0.3214911	0.3481669	0.3743283	0.4235876	0.5058927	0.5899605

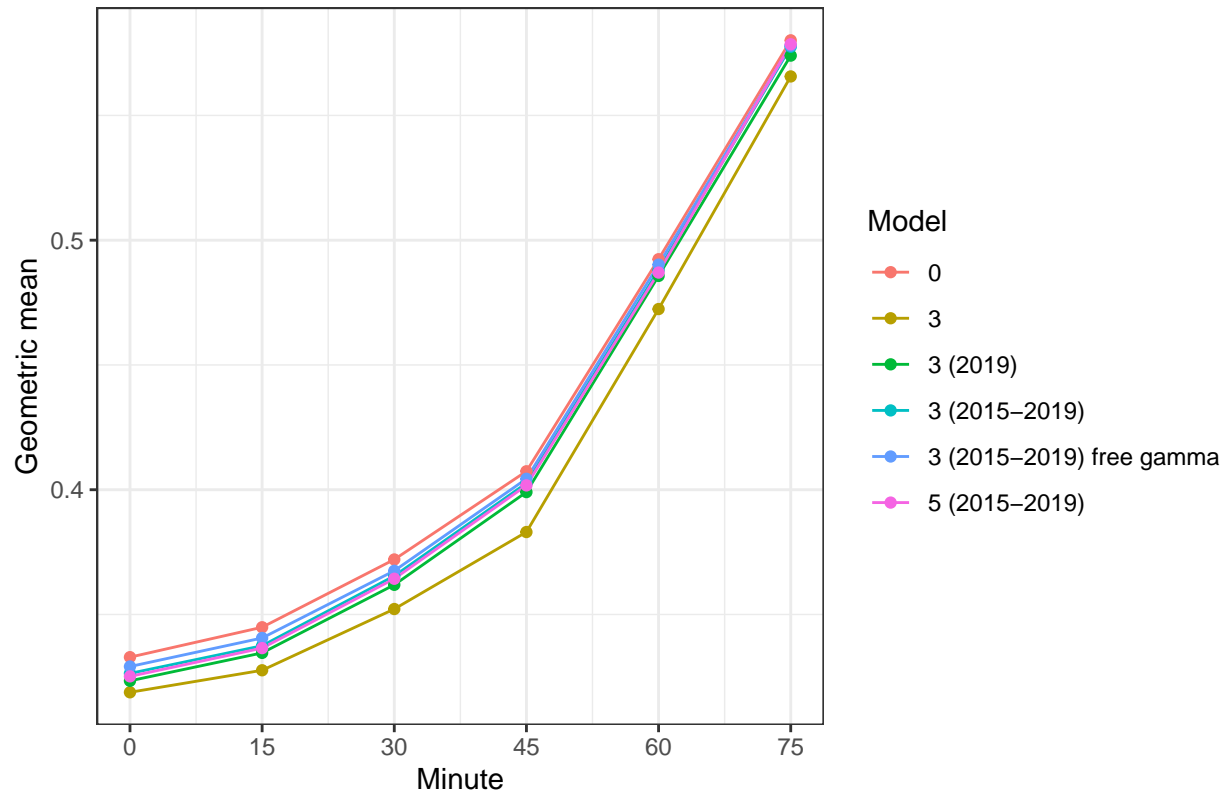| Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
5 (2015-2019)	0.3203950	0.3478352	0.3725801	0.4238746	0.5042760	0.5965536

```r
first_200 = tibble(RPS = apply(HDA[c(1:200), c(189:224)], 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 matches


```
first_200 %>%
  pivot_wider(id_cols = "Model", values_from = "RPS", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```
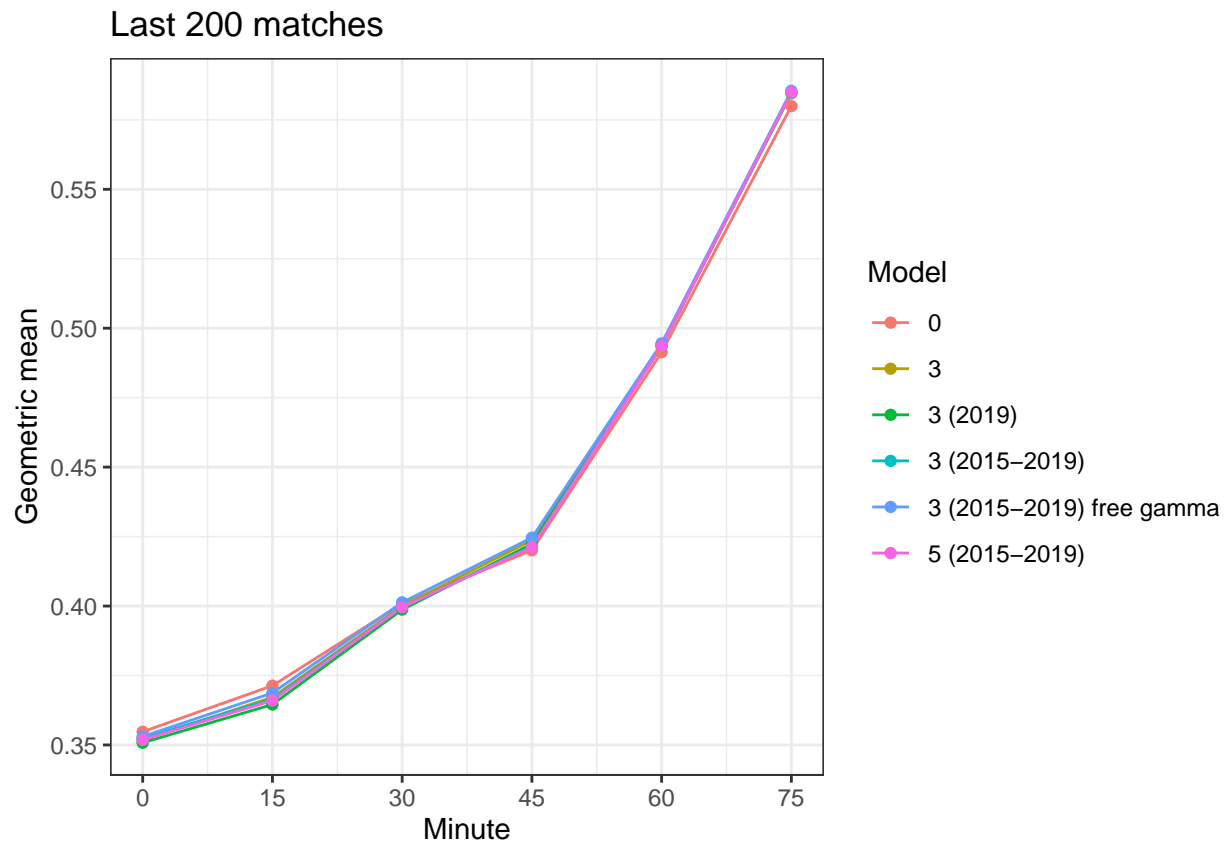
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.3328969	0.3448813	0.3720426	0.4073875	0.4923749	0.5800863
3	0.3188151	0.3276395	0.3521856	0.3830418	0.4723963	0.5656215
3 (2019)	0.3234153	0.3346044	0.3618892	0.3990470	0.4856824	0.5739646
3 (2015-2019)	0.3263835	0.3374837	0.3654488	0.4026106	0.4880504	0.5775051
3 (2015-2019) free gamma	0.3291449	0.3405805	0.3675053	0.4043682	0.4901964	0.5779762
5 (2015-2019)	0.3252040	0.3365195	0.3643146	0.4017299	0.4870210	0.5785054

```
last_200 = tibble(RPS = apply(HDA[c(134:333), c(189:224)], 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 matches

```
last_200 %>%
  pivot_wider(id_cols = "Model", values_from = "RPS", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```
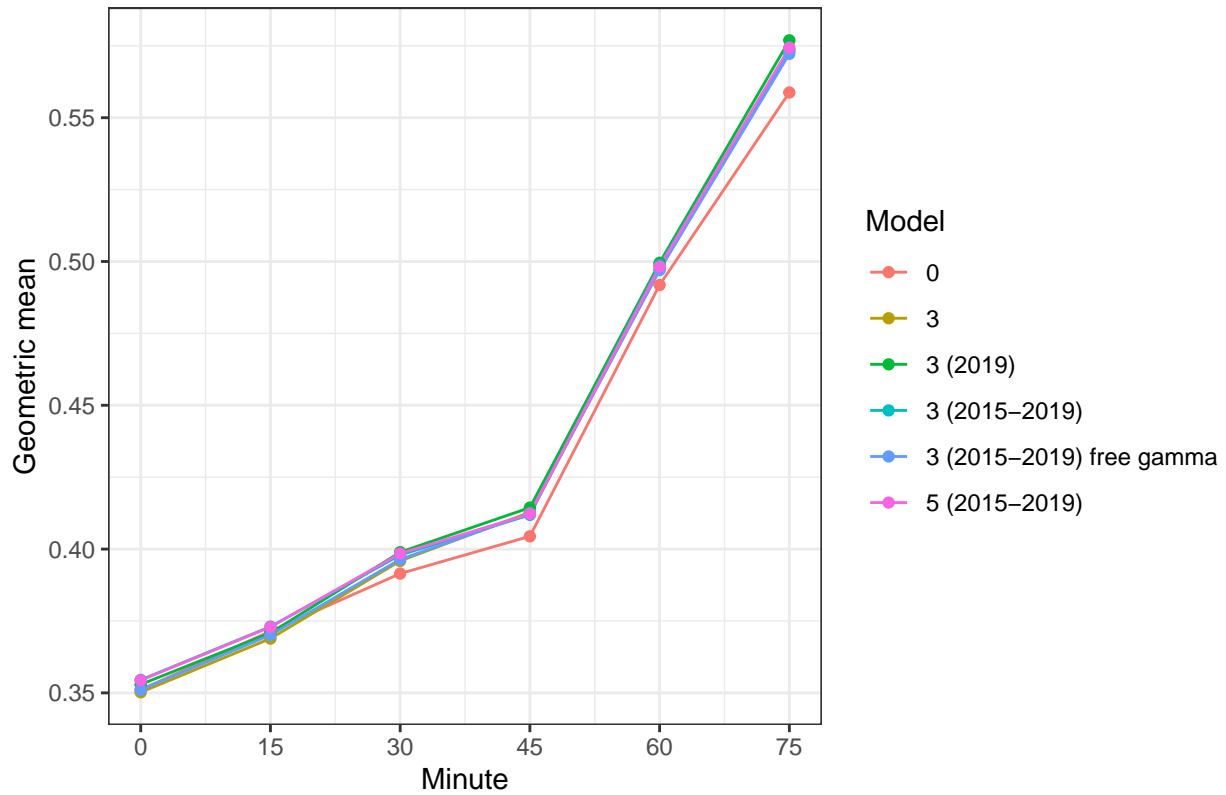
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.3547484	0.3713414	0.4008703	0.4200557	0.4913387	0.5799796
3	0.3517310	0.3671720	0.4000142	0.4236709	0.4935843	0.5845721
3 (2019)	0.3507890	0.3645648	0.3987324	0.4223301	0.4935809	0.5848193
3 (2015-2019)	0.3523944	0.3665182	0.3996135	0.4214469	0.4937446	0.5847596
3 (2015-2019) free gamma	0.3530408	0.3687099	0.4013190	0.4245910	0.4945348	0.5854619

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
5 (2015-2019)	0.3518371	0.3659931	0.3995734	0.4211897	0.4935465	0.5849290

```r
last_100 = tibble(RPS = apply(HDA[c(234:333), c(189:224)], 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 matches


```
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	0.3511211	0.3714316	0.3914752	0.4044365	0.4918286	0.5587610
3	0.3502005	0.3688780	0.3958788	0.4127854	0.4972130	0.5730454
3 (2019)	0.3529394	0.3708479	0.3989267	0.4144206	0.4995515	0.5768990
3 (2015-2019)	0.3545058	0.3730363	0.3979432	0.4118977	0.4987105	0.5736065
3 (2015-2019) free gamma	0.3510659	0.3701216	0.3964136	0.4124357	0.4970231	0.5722154
5 (2015-2019)	0.3543661	0.3729071	0.3984530	0.4122311	0.4982268	0.5743571

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

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

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

```r
all_reds = tibble(RPS = apply(HDA_reds[,c(189:224)], 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_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")
```
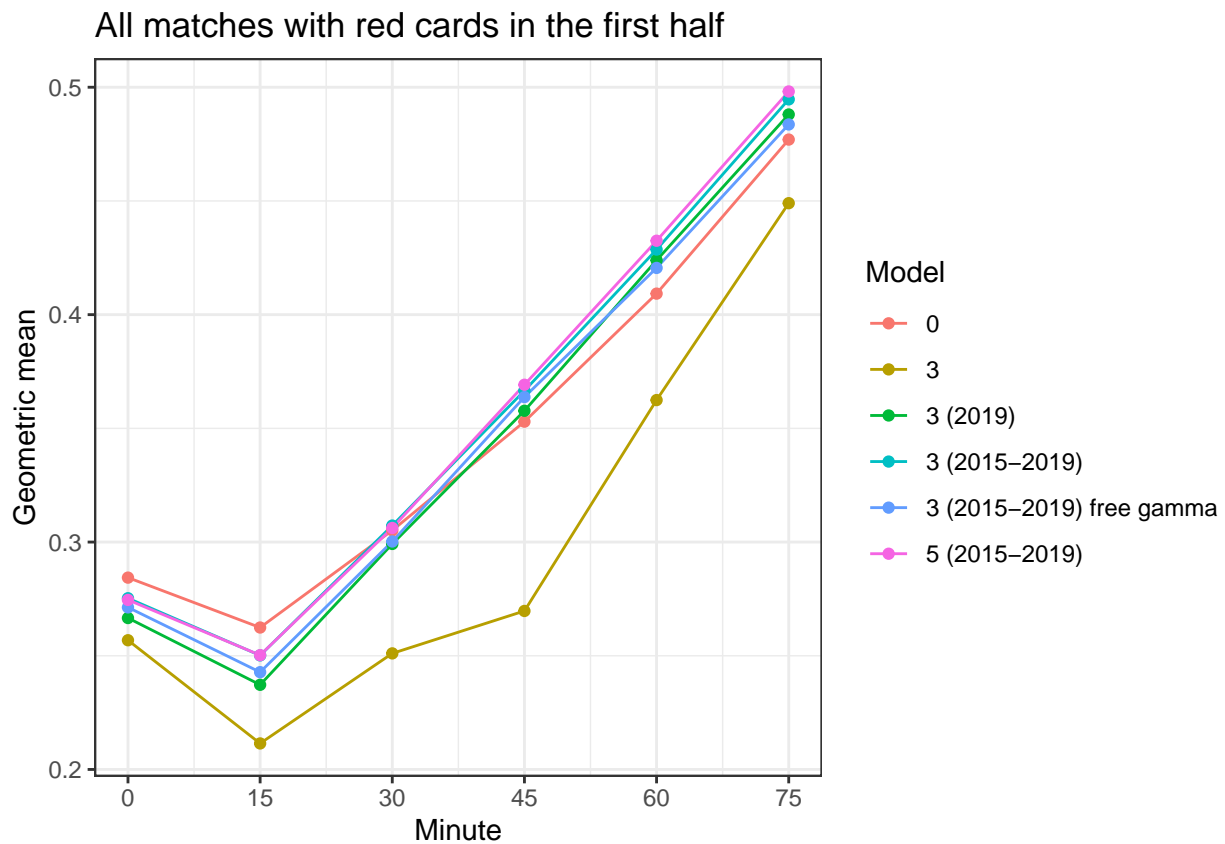


```r
all_reds %>%
  pivot_wider(id_cols = "Model", values_from = "RPS", names_from = "Minute",
```

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

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
0	0.2843631	0.2624180	0.3051243	0.3529510	0.4092512	0.4769856
3	0.2568219	0.2114438	0.2510404	0.2697029	0.3624380	0.4490213
3 (2019)	0.2665939	0.2372314	0.2991672	0.3577324	0.4239187	0.4880296
3 (2015-2019)	0.2752970	0.2501513	0.3072525	0.3665565	0.4286207	0.4946650
3 (2015-2019) free gamma	0.2712727	0.2427910	0.3002434	0.3637617	0.4205542	0.4836446
5 (2015-2019)	0.2745560	0.2502595	0.3061933	0.3691828	0.4324993	0.4981429