

Previsão do n.º de sets para conclusão de um jogo de ténis profissional

2ª semana



Progresso semanal

- Verificação da estrutura dos dados
- Tratamento dos dados relativos as colunas date e locations
- Transformação dos jogos

Verificação de estrutura de dados

- Distincts
- Conclusão sobre cada coluna

Transformação dos jogos

- *Relevância*

1.308.835 jogos

9.953 jogadores

- *Comando R:*

```
matches <- atp %>%  
  filter(Oponent!="bye") %>%  
  mutate(winner=ifelse(WL=="W", PlayerName, Oponent), loser=ifelse(WL=="L", PlayerName, Oponent)) %>%  
  group_by(Tournament, GameRound, Date, winner, loser) %>%  
  summarise()
```

682.069 matches

Colunas date

- *Período de dias de um partida*
- *12360 partidas só com 1 dia(a partida começou e acabou no mesmo dia)*
- *Análise da duração dos torneios (valores negativos?)*

tournament.duration	n	freq
<dbl>	<int>	<dbl>
6	1144856	87.4713772171
13	118095	9.0229096869
5	29228	2.2331309905
0	12445	0.9508455993
7	1672	0.1277471950
14	988	0.0754869789
20	306	0.0233795704
9	236	0.0180313026
12	231	0.0176492835
27	180	0.0137526885
11	162	0.0123774196
8	137	0.0104673240
2	108	0.0082516131
-1	63	0.0048134410
4	54	0.0041258065
3	30	0.0022921147
17	25	0.0019100956
10	9	0.0006876344
48	6	0.0004584229
-3646	4	0.0003056153

Recursos utilizados

- *Ficheiro atpplayers.json*
- *Ficheiro CountryAlias.csv*
- *Software R*

Para a semana...

- Acabar o tratamento
- Mudar para matches
- Pôr features