

Women's Football Results: Analysis and Visualization

Resultados de Competições Internacionais de Futebol Feminino por Marina Fajardo

Link: <https://www.kaggle.com/martj42/womens-international-football-results>

Dados técnicos: conjunto de dados de domínio público do Kaggle, nota 10 em “usability”, disponibilizado via arquivo csv de 52KB. Possui 6.284 views, 978 downloads e 18 notebooks.

Atualmente conta com 4.163 resultados de partidas de futebol feminino de principais torneios mundiais bem como amistosos ocorridos entre 01/11/1969 e 11/03/2020. Os dados foram raspados da Wikipedia, fifa.com, rsssf.com e alguns sites de associações individuais de futebol.

Tags: sports, football, history, internacional relations

Introdução

Após download do dataset e tratamento de dados, foram feitas análises que abordam os placares mais comuns nas competições, os times com maiores saldos de gols, além de quais foram mais participativos e como a composição do campeonato mudou ao longo dos anos.

Dicionário de Dados

Variável	Descrição	Tipo de Dados	Valores Permitidos
<i>date</i>	data da partida	IDate	datas
<i>home_team</i>	nome do time da casa	Character	texto
<i>away_team</i>	nome do time convidado	Character	texto
<i>home_score</i>	placar do time da casa (tempo normal de jogo + prorrogações. Pênaltis não considerados)	Integer	números inteiros
<i>away_score</i>	placar do time convidado (tempo normal de jogo + prorrogações. Pênaltis não considerados)	Integer	números inteiros
<i>tournament</i>	nome do torneio	Character	texto
<i>city</i>	nome da cidade onde a partida foi realizada	Character	texto
<i>country</i>	nome do país onde a partida foi realizada	Character	texto
<i>neutral</i>	VERDADEIRO/FALSO - indica se a partida foi realizada em um local neutro para as duas equipes	Logical	valores lógicos

Preparação dos Dados

Verificando e acertando nomes escritos de formas diferentes:

Table 1: Amostra do Dataset

date	home_team	away_team	home_score	away_score	tournament	city	country	neutral
1969-11-01	Italy	France	1	0	Euro	Novara	Italy	FALSE
1969-11-01	Denmark	England	4	3	Euro	Aosta	Italy	TRUE
1969-11-02	England	France	2	0	Euro	Turin	Italy	TRUE
1969-11-02	Italy	Denmark	3	1	Euro	Turin	Italy	FALSE
1975-08-25	Thailand	Australia	3	2	AFC Championship	Hong Kong	Hong Kong	TRUE
1975-08-25	Hong Kong	New Zealand	0	2	AFC Championship	Hong Kong	Hong Kong	FALSE

Note:

Womens International Football Results (1969 - 2020)

Alterações realizadas:

Anterior	Novo
Curaçao	Curaçao
DR Congo e Congo	Congo
India, India N e India S	India
Réunion	Reunion
São Tomé and Príncipe	San-Tome and Principe
Copa América	Copa América
Euro	UEFA Euro

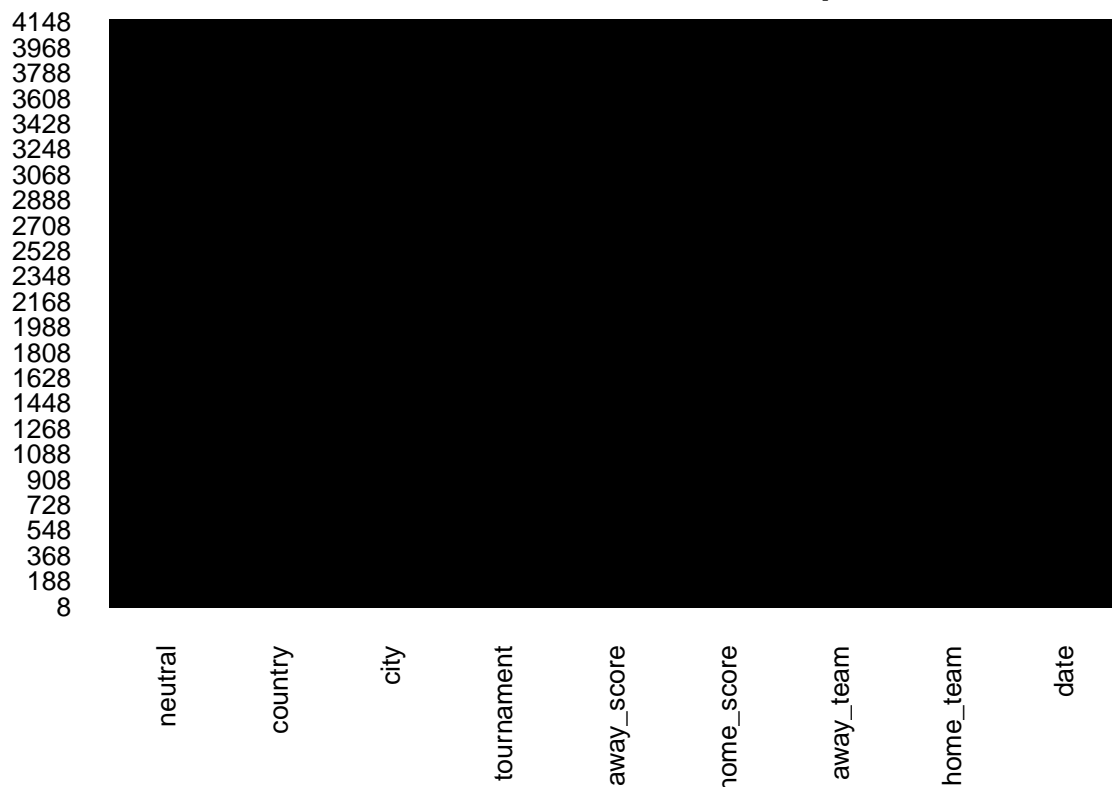
Há, ainda, alguns países que não sofreram alterações pois eram antigas uniões entre terras, que eram verdadeiras de acordo com as datas dos jogos. São eles: Czechoslovakia (recentes Czech Republic e Slovakia), Netherlands Antilles (conjunto de terras que originou Netherlands) e Serbia and Montenegro (recentes Serbia e Montenegro)

A coluna que contém data está classificada como “IDate” e “Date”. Contudo, caso seja necessário no futuro realizar algum outro tipo de análise utilizando hora, seria conveniente transformar a coluna “Date” para o padrão POSIXct.

As colunas que contém os times, países, torneios e cidades (home_team, away_team, country, tournament e city) devem ser convertidas para fatores.

Nenhum valor NA foi encontrado no dataset

Womens International Football Results – Mapa de Dados Missir



Análise Exploratória

Criação de um dataframe completo para análises. Colunas inseridas:

Coluna	Descrição
<i>Placar</i>	Juntação de home_score e away_score, indicando o placar da partida
<i>Vencedor</i>	Informa o vencedor da partida
<i>Ano</i>	Informa o ano de realização da partida, útil para plotagens futuras
<i>Década</i>	As décadas foram divididas em 4 com 10 anos cada e 1, a primeira, contendo 11 anos. Década 1 = 1969 - 1980, Década 2 = 1981 - 1990, Década 3 = 1991 - 2000, Década 4 = 2001 - 2010 e Década 5 = 2011 - 2020

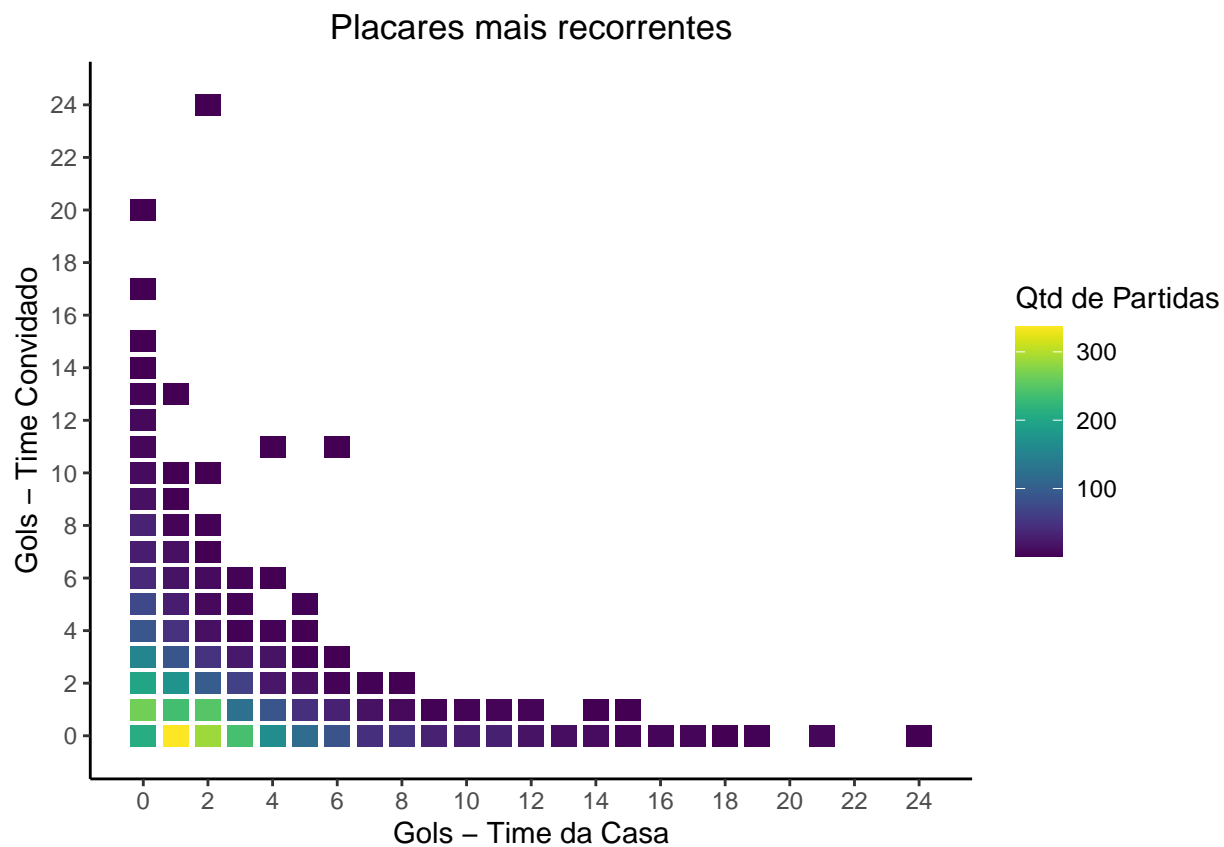
Novo data frame de trabalho:

date	home_team	away_team	home_score	away_score	tournament	city
1969-11-01	Italy	France	1	0	UEFA Euro	Novara
1969-11-01	Denmark	England	4	3	UEFA Euro	Aosta
1969-11-02	England	France	2	0	UEFA Euro	Turin
1969-11-02	Italy	Denmark	3	1	UEFA Euro	Turin
1975-08-25	Thailand	Australia	3	2	AFC Championship	Hong Kong
1975-08-25	Hong Kong	New Zealand	0	2	AFC Championship	Hong Kong

country	neutral	placar	vencedor	ano	decada
Italy	FALSE	1 x 0	Italy	1969	Década 1
Italy	TRUE	4 x 3	Denmark	1969	Década 1
Italy	TRUE	2 x 0	England	1969	Década 1
Italy	FALSE	3 x 1	Italy	1969	Década 1
Hong Kong	TRUE	3 x 2	Thailand	1975	Década 1
Hong Kong	FALSE	0 x 2	New Zealand	1975	Década 1

Analisando Placares

Abaixo um mapa de calor que demonstra os placares que mais ocorreram ao longo do campeonato. Vemos claramente que o placar de 1 x 0 foi a combinação mais recorrente.

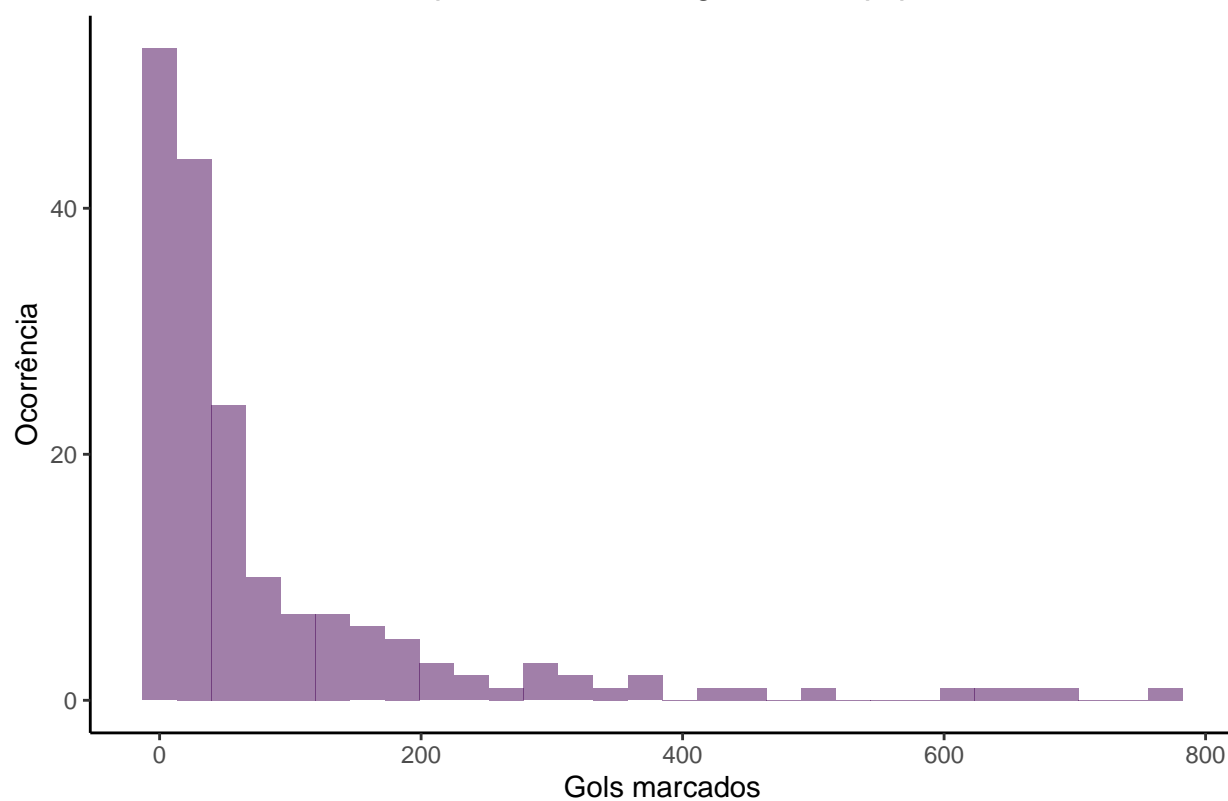


Gols marcados, sofridos e saldo de gols

Ranking times que mais fizeram gols

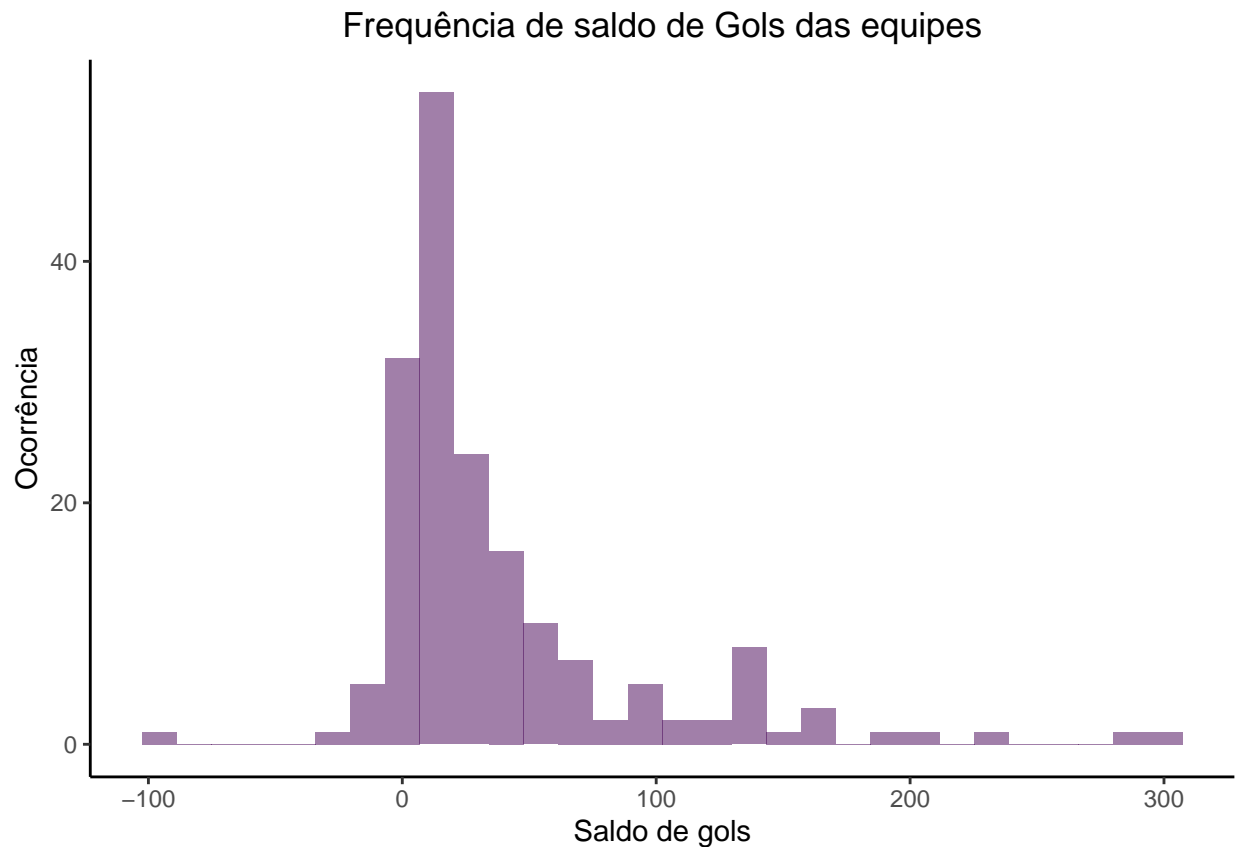
Equipe	Gols feitos
Germany	769
Norway	689
United States	651
China PR	637
Sweden	611
Denmark	495
Japan	460
Brazil	417
England	384
New Zealand	368

Frequência totais de gols das equipes



Ranking times com melhores saldos de gols (gols feitos - gols sofridos)

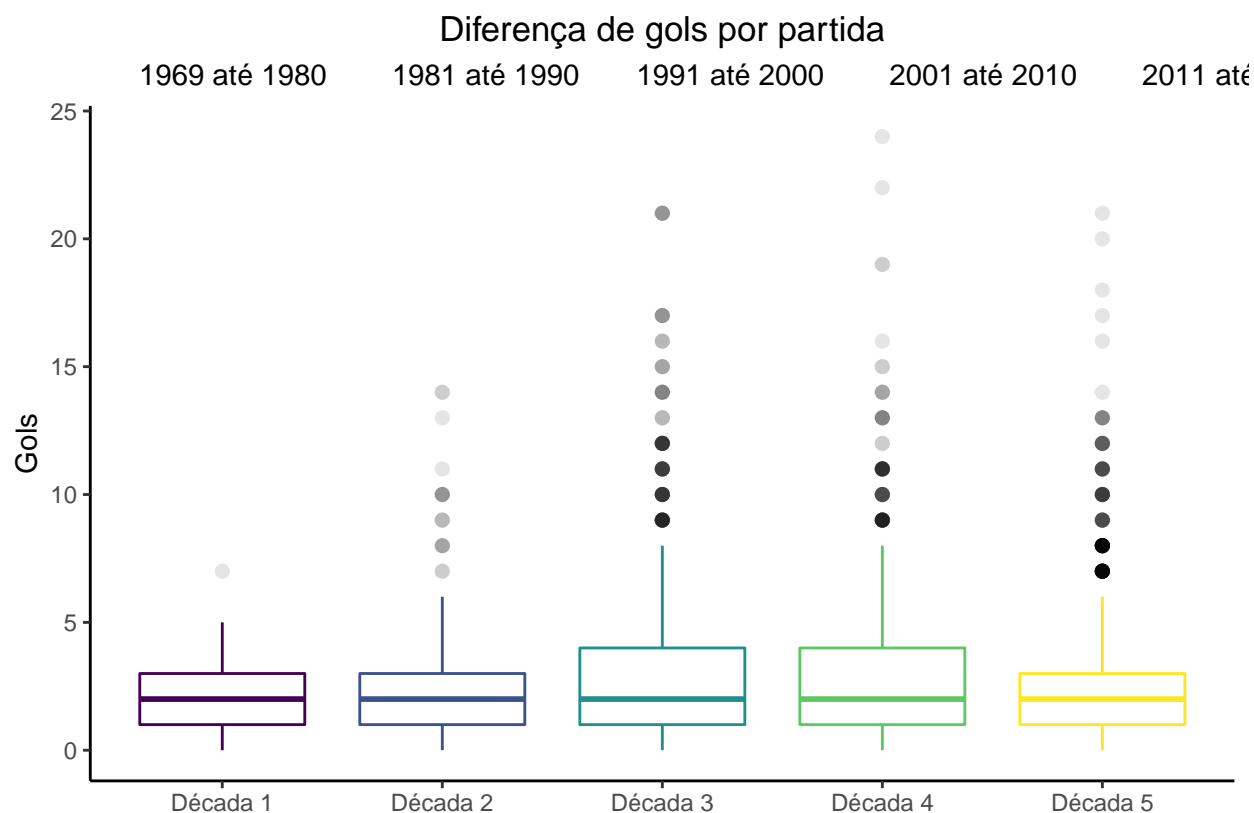
Equipe	Saldo de Gols
China PR	296
United States	292
Japan	238
Norway	210
Trinidad and Tobago	193
Canada	169
North Korea	158
Finland	158
Brazil	153
Germany	143



Apesar da Alemanha ter potencial ofensivo expressivo (tendo marcado 769 gols), ela ocupa apenas a décima posição se analisado o saldo de gols (com 143 gols). China, Estados Unidos e Japão aparentam ter potencial ofensivo e defensivo mais equilibrado. Existem muitas equipes com saldo de gols próximos de zero e algumas ainda com saldos negativos.

Ainda podemos observar pela análise dos histogramas, a maior frequência de equipes que não marcam muitos gols ou não possuem saldos de gols tão elevados. As poucas super goleadoras ou com alto saldo de gols podem ser consideradas outliers.

Ainda podemos verificar a diferença de gols entre as partidas e se houve muita mudança ao longo dos anos: Entende-se por diferença de gols a subtração dos gols feitos pelo time da casa dos gols feitos pelo time convidado.



Nota-se que não há muitas alterações relevantes se compararmos as médias de diferenças de gols entre as décadas. No entanto, podemos perceber que a partir de 1991 houve importante aumento de maiores diferenças, o que pode indicar certa disparidade na qualidade de algumas equipes.

Vitórias e Derrotas

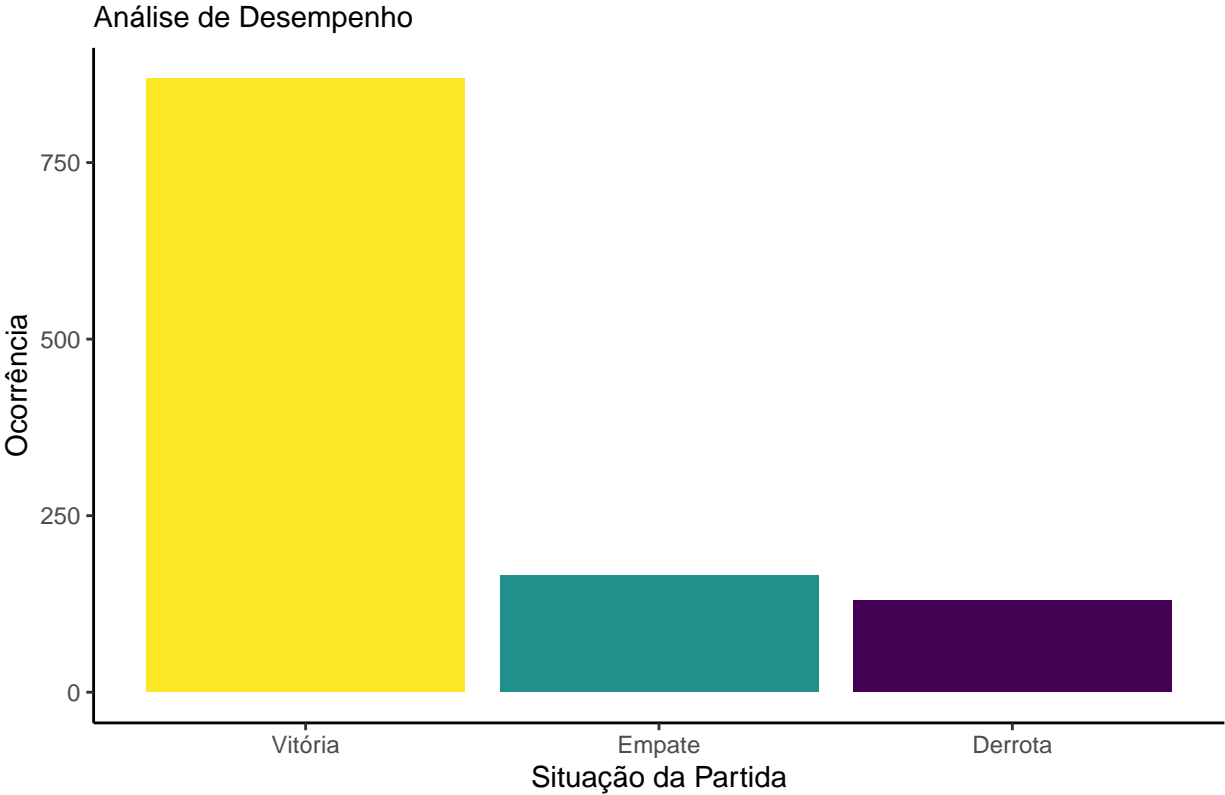
Abaixo os times mais e menos vitoriosos

Equipe	Quantidade de Vitórias
Germany	190
Norway	184
United States	183
Sweden	165
China PR	147
Samoa	1
Serbia and Montenegro	1
Tahiti	1
Tajikistan	1
Uganda	1

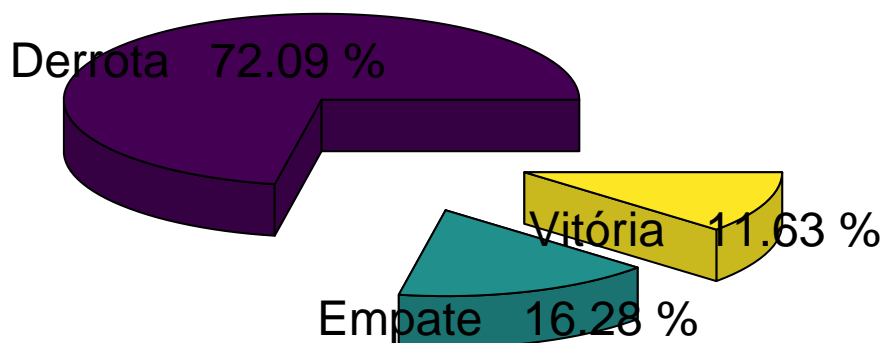
Alemanha, Noruega e Estados Unidos são as equipes mais vitoriosas, enquanto Tahiti, Tadjiquistão e Uganda compõem um grande grupo que teve apenas uma vitória ao longo de sua participação.

Podemos, também, plotar um balanço de desempenho das equipes mais e menos vitoriosas:

Top 5 Mais Vitoriosos



Análise de Desempenho Top 5 menos vitoriosos

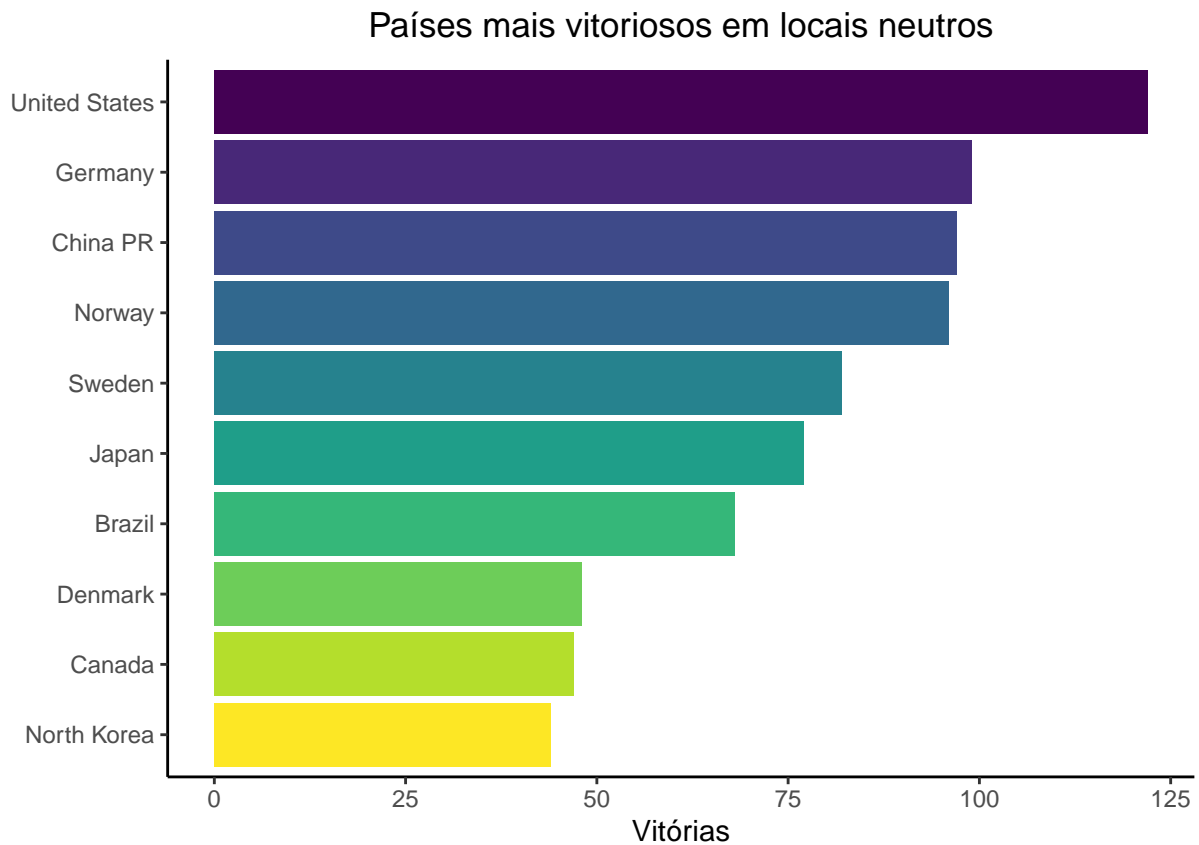


Enquanto as equipes mais vitoriosas obtiveram, em sua esmagadora maioria, vitórias em suas partidas, as menos vitoriosas apresentaram menos de 12% de jogos bem sucedidos.

Se analisarmos as competições que não ocorreram em lugares neutros, a Noruega lidera o ranking de vitórias, seguida pela Alemanha e Dinamarca. Os Estados Unidos não aparecem entre os dez primeiros lugares. Jogar em casa pode ser um fator relevante para a equipe estadunidense.

Equipe	Vitórias fora de casa
Norway	44
Germany	42
Denmark	39
Sweden	39
England	35
Italy	35
Russia	28
France	26
Iceland	22
Netherlands	21

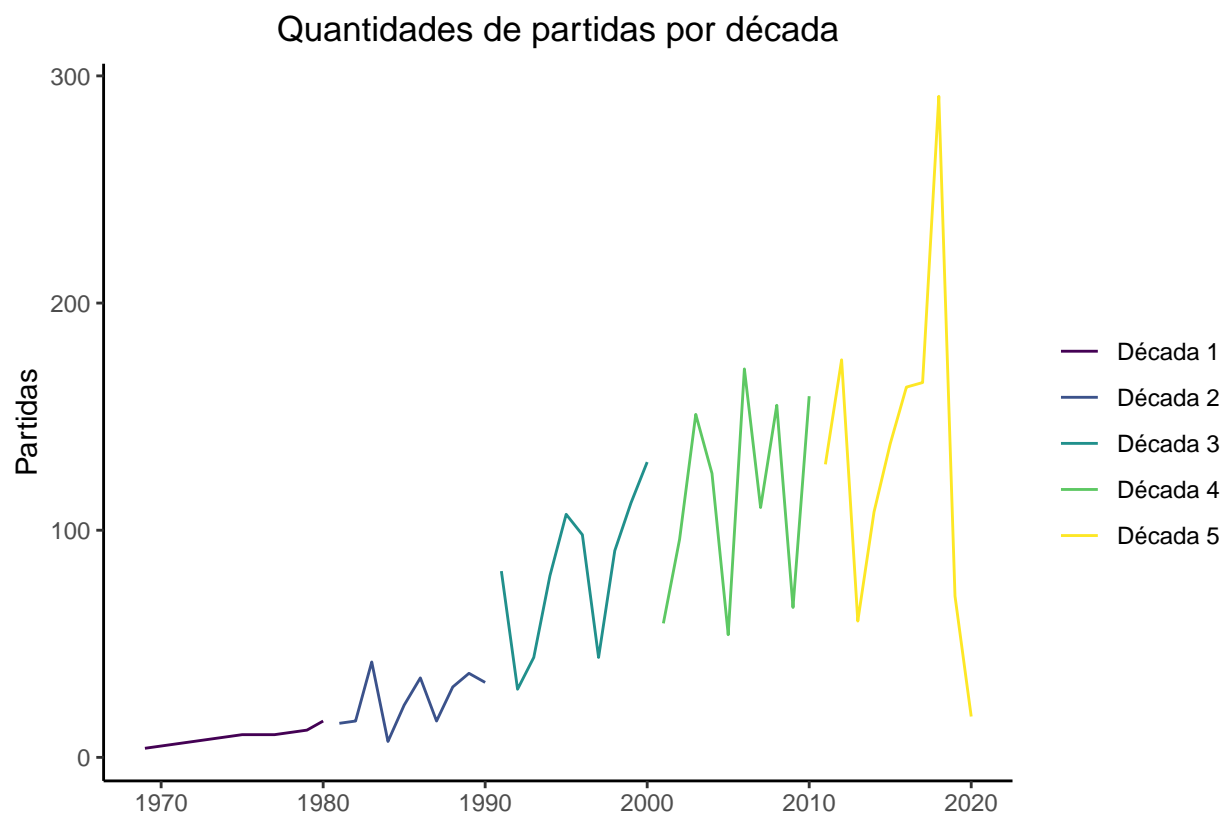
Apesar das poucas vitórias, Portugal foi o local que mais sediou competições:



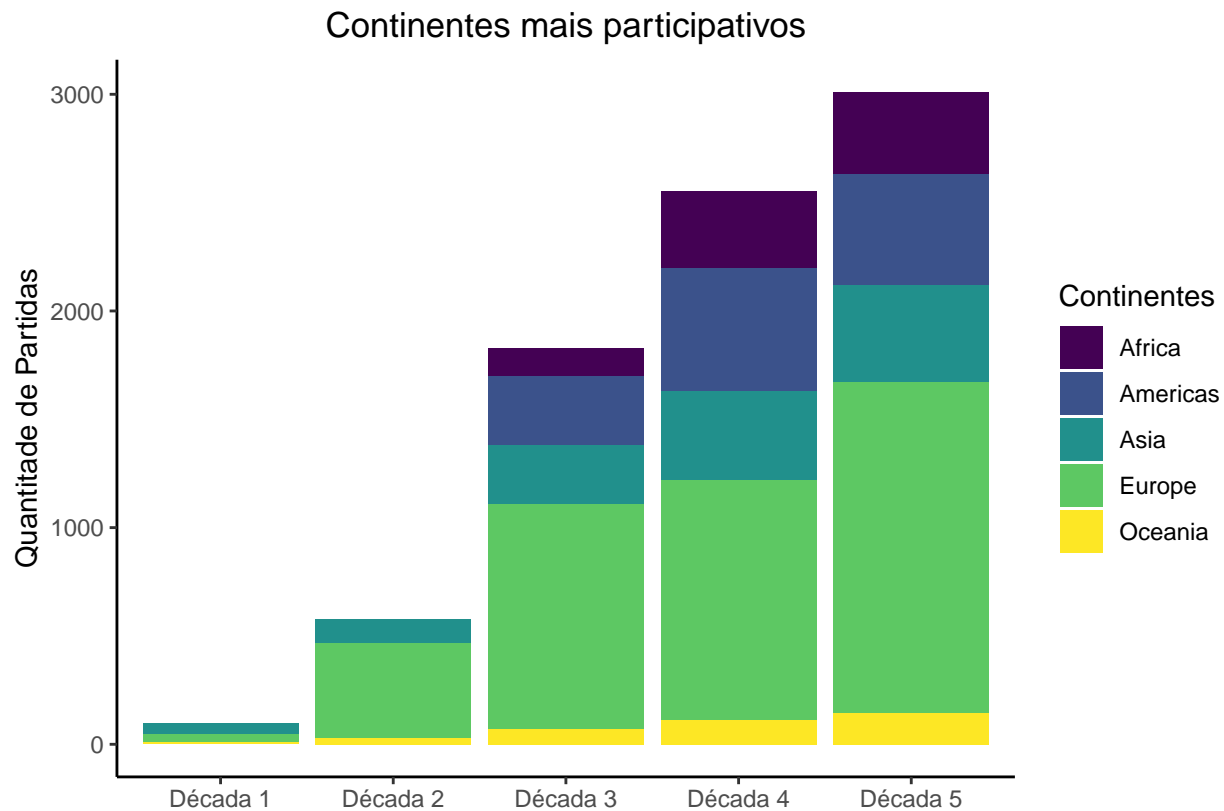
Alemanha, China e Noruega voltam a aparecer como importantes equipes. Os Estados Unidos demonstram que possuem uma equipe consistente, tendo além de um ótimo saldo de gols e sustentando-se como uma das equipes mais vitoriosas, capacidade de realizar bons jogos mesmo fora de casa.

Times e países mais participativos:

Percebemos no gráfico abaixo o aumento significativo da quantidade de partidas realizadas ao longo das décadas



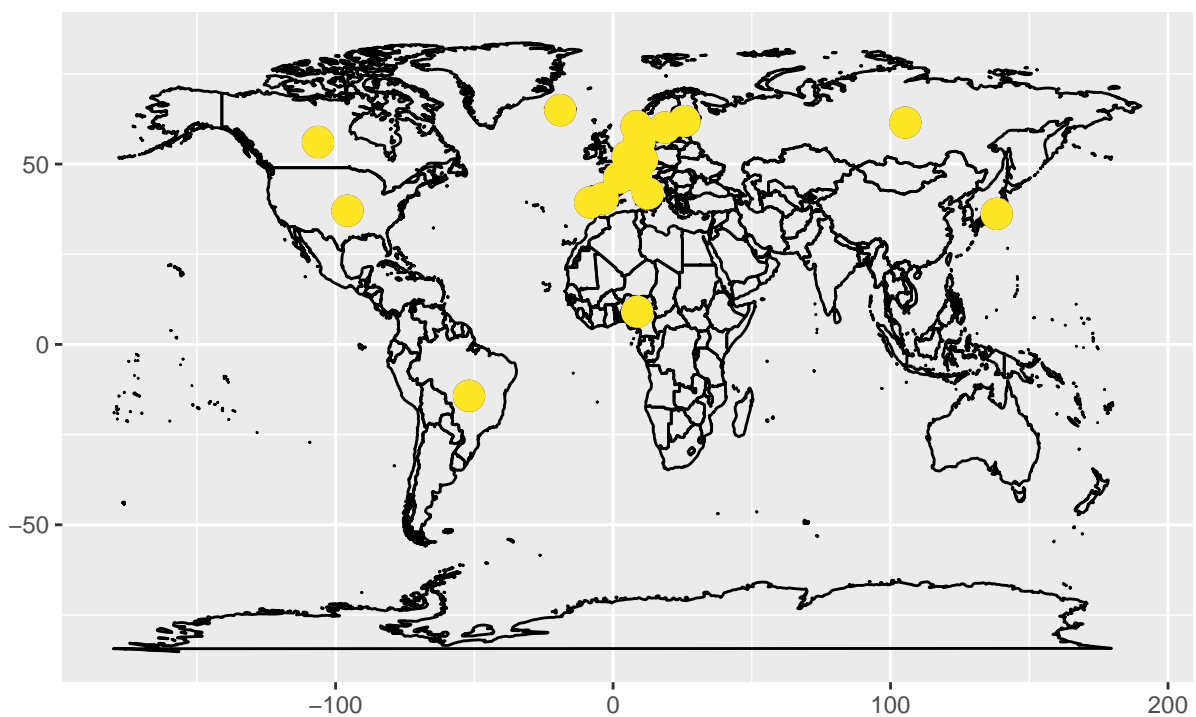
Quais são os continentes mais participativos e como mudou a composição dos campeonatos de forma geral?



No início dos anos 70, basicamente países pertencentes à Europa, Ásia e Oceania disputavam torneios internacionais. A partir de 1991, percebe-se entrada de países do continente africano. Mesmo assim, em 2020, apesar da forte presença da América, a Europa segue sendo o mais relevante continente na composição dos campeonatos mundiais de futebol feminino.

De forma geral, quais foram os países mais participativos?

Países mais participativos



Observa-se claramente a hegemonia dos países europeus nos campeonatos.