

Few facts about world cups

Przez wzgląd na ubiegłoroczne mistrzostwa świata w piłce nożnej w Katarze postanowiłem dokonać ogólnego przeglądu statystyk, które mnie najbardziej interesowały w poszczególnych mundialach od 1930 do 2022 roku.

Pytania na które szukałem odpowiedzi:

1. Ile rund było w poszczególnych turniejach mistrzostw świata oraz jak na przestrzeni lat zmieniała się średnia liczba goli na mecz?
2. Czy to, że dane państwo jest organizatorem imprezy i ma po swojej stronie kibiców wpływa na wynik w danych mistrzostwach?
3. Które kontynenty dominują, jeśli chodzi o wygrane mecze oraz o całe mistrzostwa?
4. Które państwa grają najbardziej zachowawczo i miały najwięcej remisowych wyników w fazie grupowej?

Hipotezy:

1. Popularność piłki nożnej bezsprzecznie wzrosła przez stulecie, więc liczba rund oraz zespołów powinna być większa. Liczba bramek natomiast jest odwrotnie proporcjonalna liczby zespołów biorących udział w mistrzostwach, ponieważ wytrenowanie piłkarzy na wszystkich pozycjach oraz poziom sędziowania osiągnęły wyższy pułap.
2. Tak, ma to wpływ.
3. Europa i Ameryka Południowa, ponieważ kultura gry w piłkę nożną jest tam najbardziej rozwinięta.
4. Przez ostatnie dziesięciolecie Włoska drużyna była tą która grała bardzo defensywny i zachowawczy futbol.

Analiza danych

Dane mistrzostw świata 1930 – 2014 pobrałem z następującego źródła:

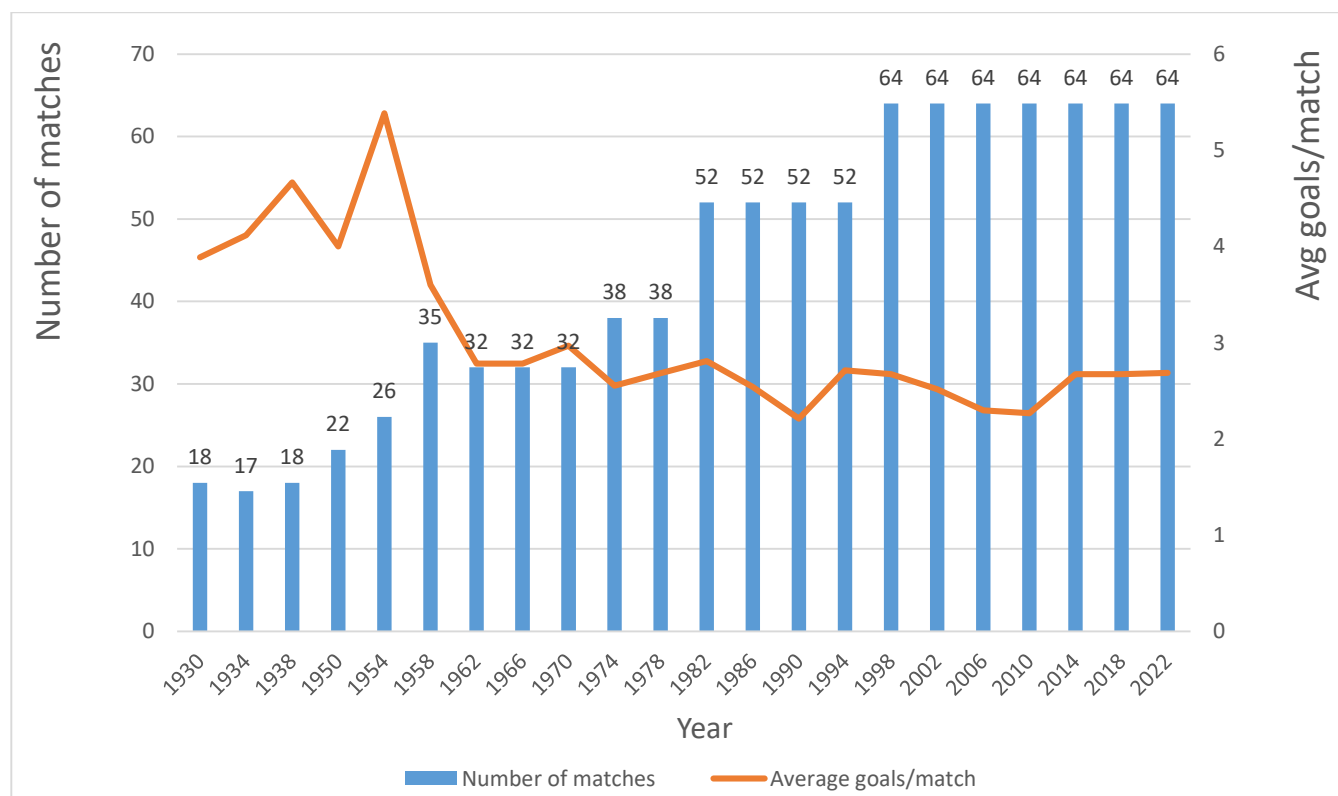
<https://data.world/sportsvizsunday/sports-viz-sundays-2018/workspace/file?filename=World+Cup+Results.xlsx>

Następnie uzupełniłem je o mundial w Rosji 2018 oraz Katarze 2022. Wyczyściłem dane, sformatowałem komórki, usunąłem zbędne oraz dodałem potrzebne kolumny do postawionych wcześniej pytań. Zbiór danych world_cup.csv zawierał ostatecznie kolumny: Year, Date, Time, Round, Stadium, City, Country, HomeTeam, HomeGoals, AwayGoals, AwayTeam, Winner, HowEnd

Drugim zbiorem danych był plik csv zawierający państwa oraz kontynenty, w których te kraje się znajdują. Ten zbiór danych wykorzystałem przy odpowiedzi na 3. pytanie. Źródło danych: https://github.com/dbouquin/IS_608/blob/master/NanosatDB_munging/Countries-Continents.csv Sformatowałem komórki oraz dodałem państwa, które już nie istnieją (np. Jugosławia) a w swojej historii brały udział w mistrzostwach świata. Zbiór danych continents.csv zawierał ostatecznie kolumny: Continent, Country.

Do analizy danych wykorzystałem SQL. Zimportowałem oba zbiory danych i wykorzystując odpowiednie zapytania tworzyłem mniejsze podzbiory, które wykorzystałem przy tworzeniu wykresów i analizie.

Na poniższym wykresie zestawilem liczbę rund rozegranych podczas poszczególnych mistrzostw świata oraz średnią liczbę strzelonych goli w meczu.



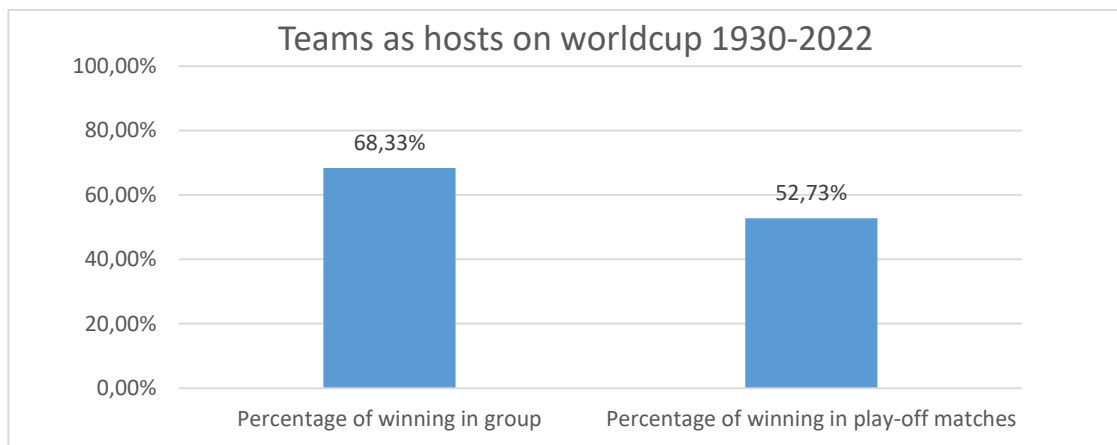
Rys. 1. Liczba rund oraz średnia liczba goli w mistrzostwach świata

Obserwując wartości poszczególnych kolumn (liczba meczy) oraz utworzonej krzywej (średnia liczba goli w meczu) można dostrzec odwrotną proporcjonalność tych dwóch zmiennych. Przy największej liczbie meczy (64 – od 1998 do 2022 roku) średnia liczba goli na mecz wynosi ok 2,5, natomiast na sześciu pierwszych mundialach (od 1930 do 1954), dla których liczba meczy nie przekroczyła 30, średnia liczba strzelonych bramek to ok 4,4. Na tak dużą różnicę mogą wpływać takie czynniki jak: większa intensywność meczów co wymusza na zawodnikach lepsze rozłożenie sił na cały turniej; poziom sędziowania oraz zasady, które na przestrzeni lat uległy zmianie; zaawansowanie piłkarzy – piłka nożna stała się bardziej popularna przez co jej poziom stale rośnie, próg wejścia jest wyższy a poziom (między pozycjami i drużynami) bardziej wyrównany.

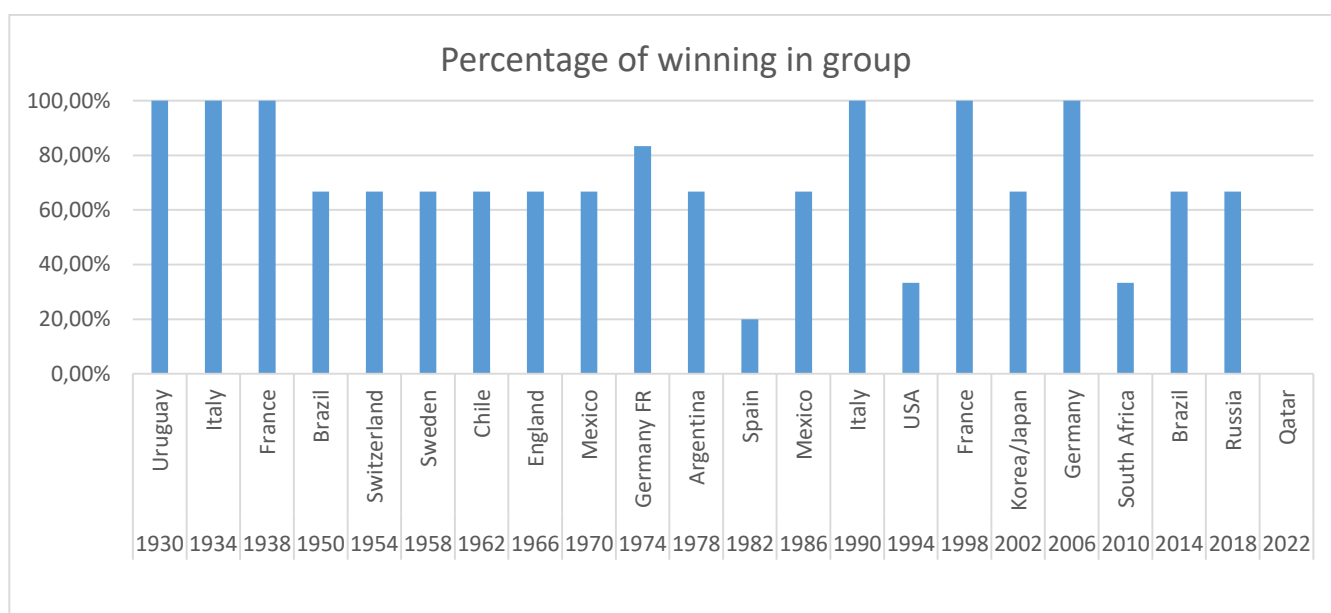
Na kolejnych wykresach został przedstawiony wpływ bycia gospodarzem i gościem mistrzostw świata na wyniki: w grupie, w fazie pucharowej, oraz zdobyte miejsca na podium. Jak się okazuje – miano gospodarza ma duży wpływ na wynik. Chciałoby się wierzyć i powiedzieć, że jest on jedynie sportowy, ale np. opinie po mundialu 2002 w Korei Południowej i Japonii nie są tak jednoznaczne. Wielu europejskich dziennikarzy oraz sportowców doszukiwało się tam stronniczości sędziów oraz korupcji.

Źródło:

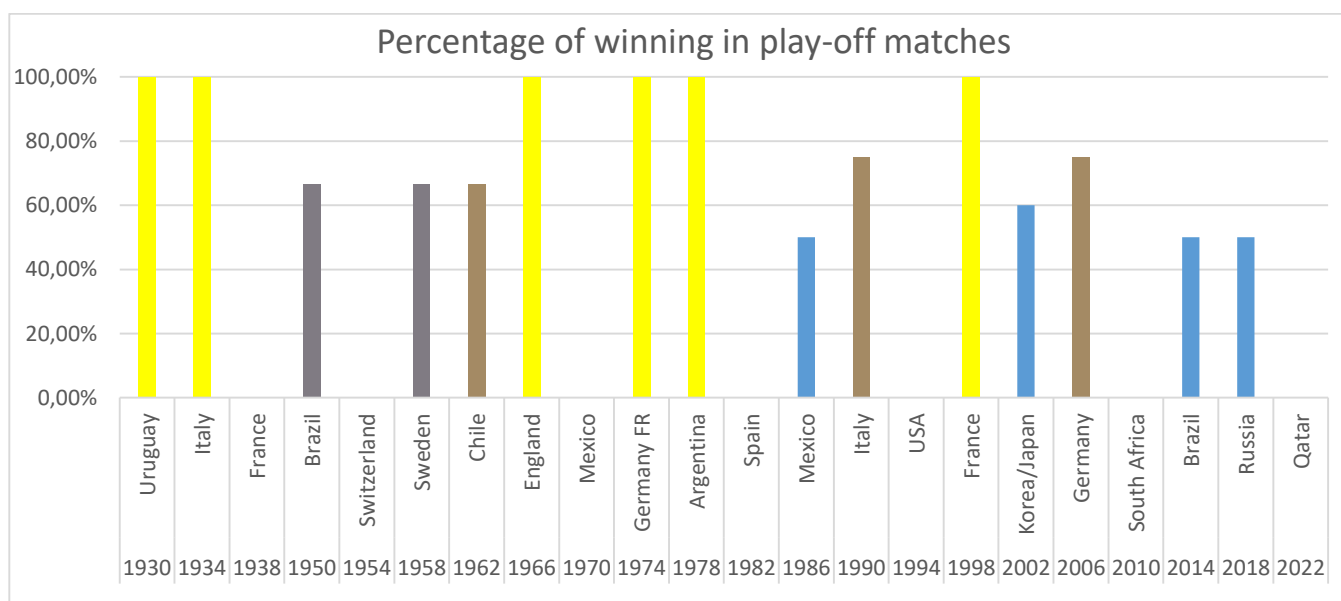
<https://www.vice.com/en/article/ywgx4y/how-the-2002-world-cup-became-the-most-controversial-tournament-in-recent-memory>



Rys. 2. Procentowy rozkład wygranych meczów przez gospodarzy mistrzostw świata w fazie grupowej oraz pucharowej



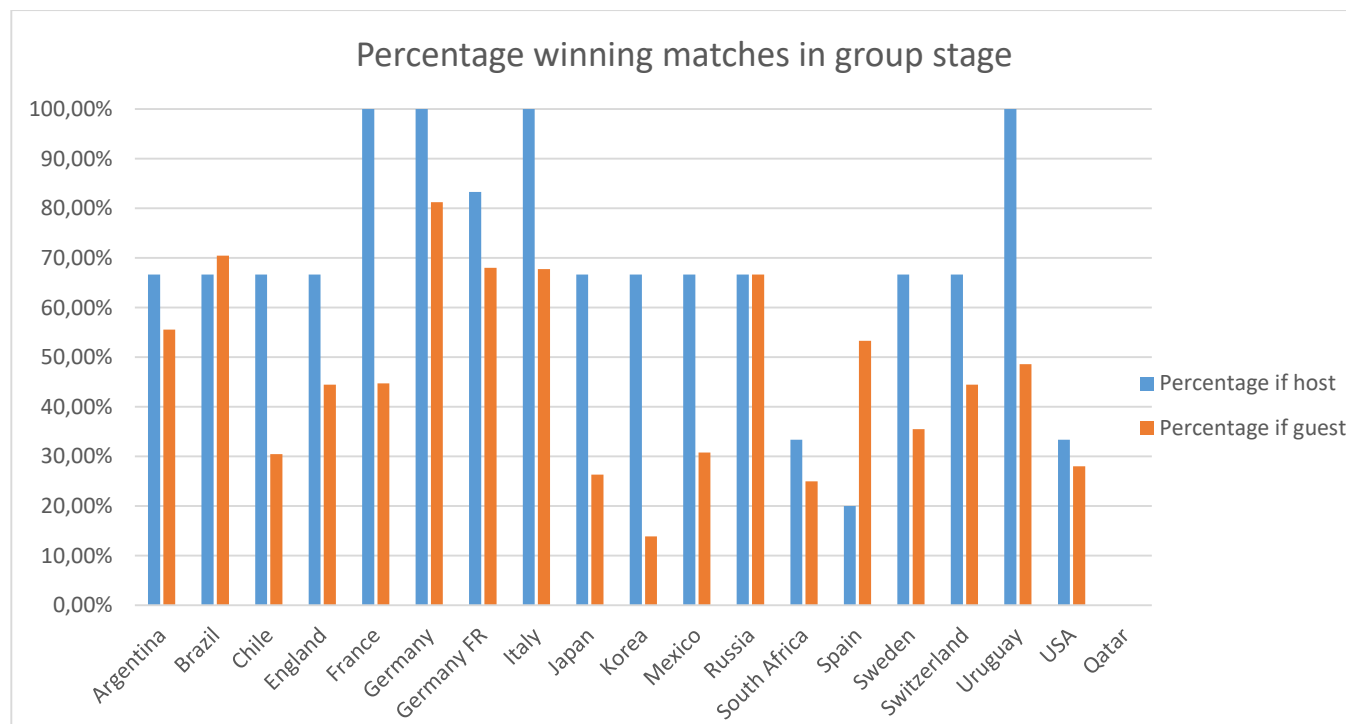
Rys. 3. Procentowy rozkład wygranych meczów przez kolejnych gospodarzy mistrzostw świata w fazie grupowej



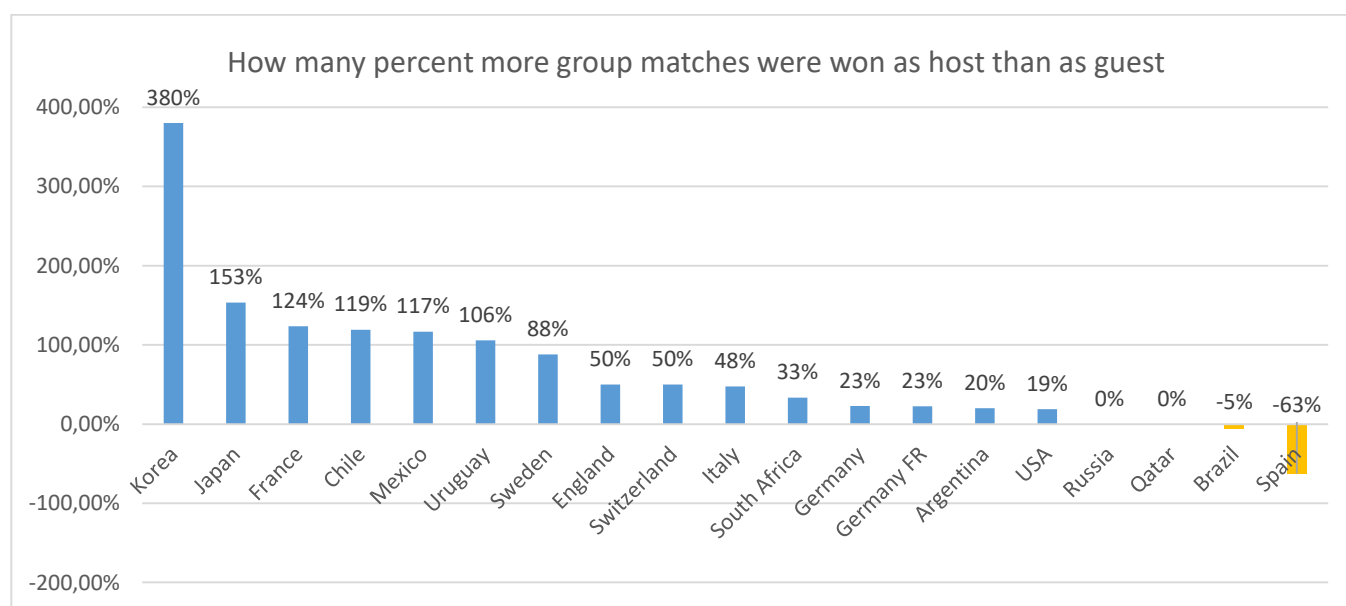
Rys. 4. Procentowy rozkład wygranych meczów przez kolejnych gospodarzy mistrzostw świata w fazie pucharowej

Przy uwzględnieniu tego, że w fazie grupowej mecze mogą zakończyć się remisem - wartość wygranych meczów na poziomie 68,33 % jest wynikiem, który świadczy o tym jak duży wpływ ma granie na „własnej ziemi” i w obecności swoich kibiców. Jedynym mundialem, na którym drużyna gospodarzy nie zwyciężyła ani jednego meczu, jest ubiegłoroczny turniej w Katarze.

Większość drużyn, które były gospodarzami mistrzostw świata zagrało w fazie pucharowej – nie zagrały w niej jedynie Hiszpania w 1982, USA w 1994 oraz wcześniej wspomniany Katar w 2022. 12 z 21 zespołów będących gospodarzami zajęło miejsce na podium (6 złotych medali – Urugwaj 1930, Włochy 1934, Anglia 1966, Niemcy Zachodnie 1974, Argentyna 1978 oraz Francja 1998; 2 srebrne – Brazylia 1950, Szwecja 1958; 3 brązowe – Włochy 1990, Korea 2002, Niemcy 2006).

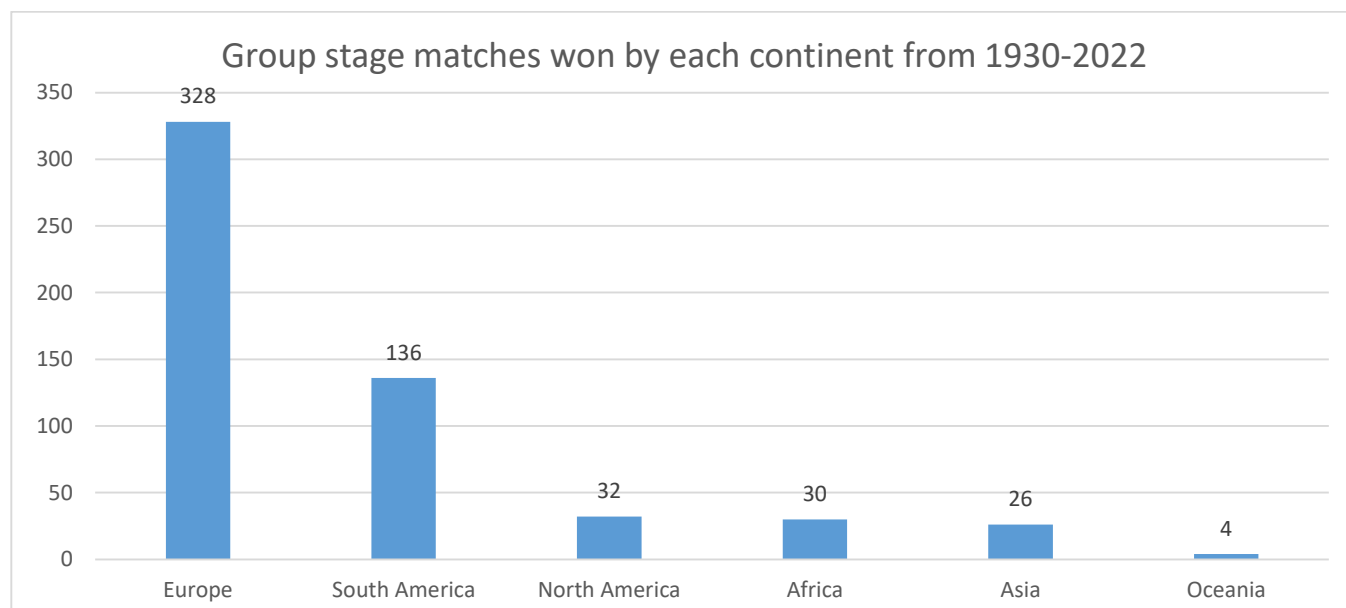


Rys. 5. Porównanie wygranych meczów w fazie grupowej jako gospodarze oraz jako goście

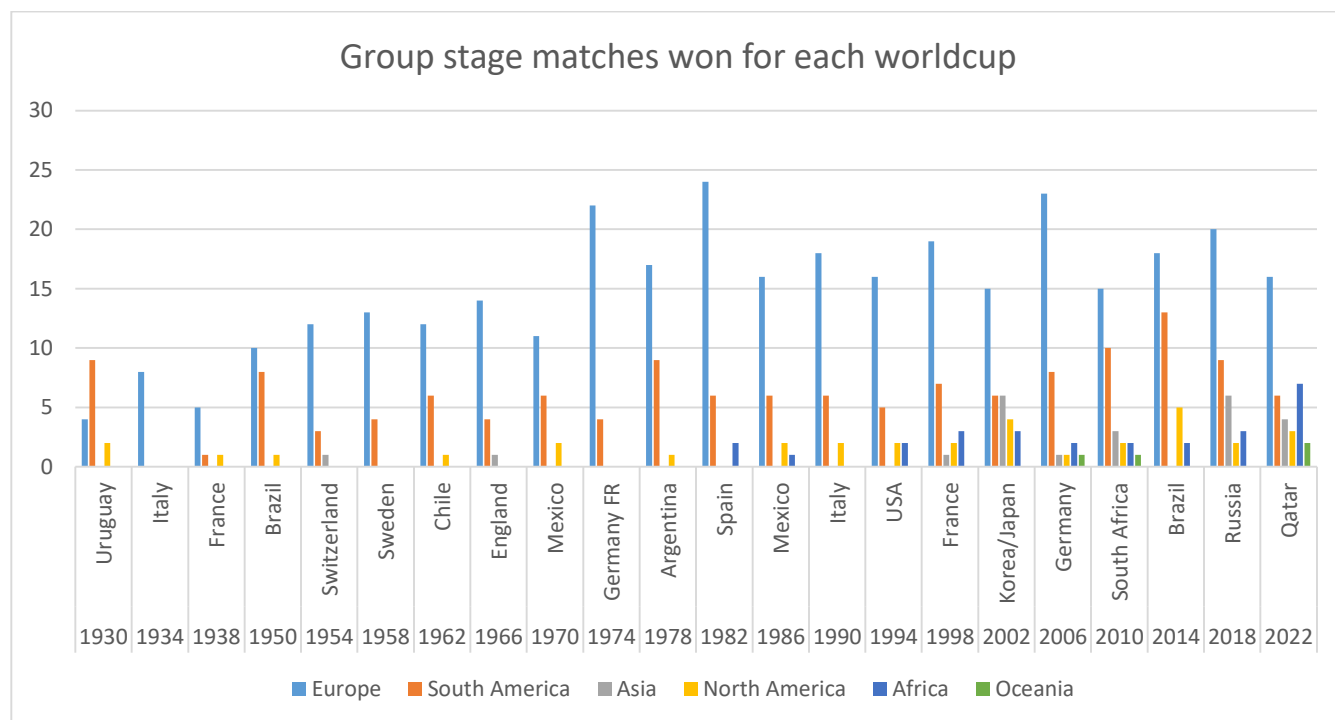


Rys. 6. O ile procent więcej meczów grupowych dana drużyna wygrała jako gospodarz niż jako gość

Powyższy wykres (Rys. 6.) może być niejako dowodem na argumentowanie tezy, która została postawiona przez wielu dziennikarzy w 2002. Dwa państwa o najwyższej różnicy wygranych meczów grupowych jako gospodarz oraz jako gość, to gospodarze pierwszego mundialu w XXI w. Analizując te wykresy można dostrzec, że granie przed własną publicznością naprawdę sprzyja dużej większości drużyn. Dwie drużyny (Rosja i Katar) mają identyczny stosunek liczby meczów wygranych w roli gospodarza oraz gościa, przy czym mundial w 2022 roku był pierwszym turniejem najwyższej rangi dla Kataru. Ze wszystkich gospodarzy są tylko dwie drużyny które notują niższy procent zwycięstw przed własną publicznością. W przypadku Hiszpanii jest to wynik ok 63 %.



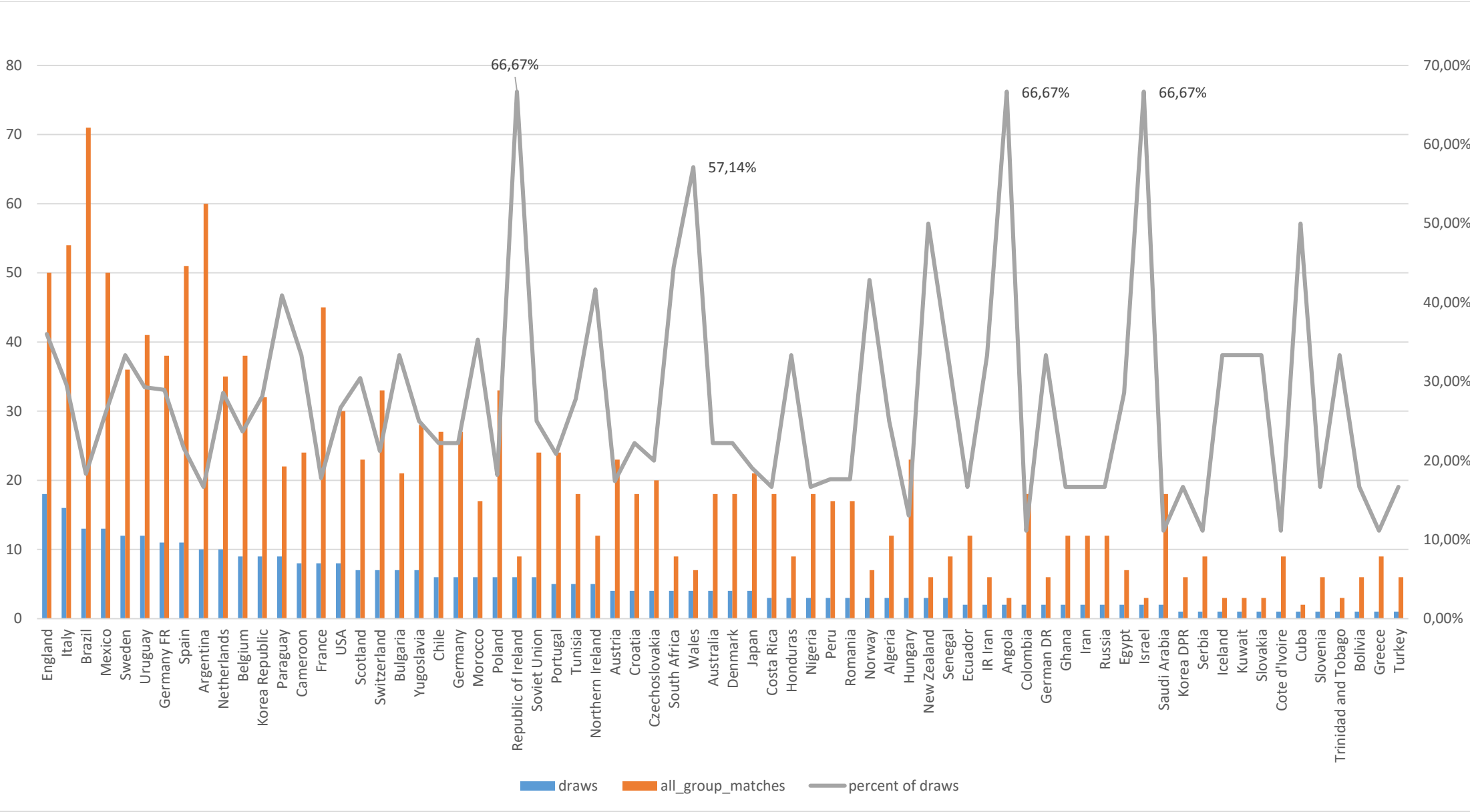
Rys. 7. Liczba wygranych meczów grupowych przez każdy kontynent od 1930 do 2022



Rys. 8. Liczba wygranych meczów grupowych przez każdy kontynent na poszczególnych mistrzostwach

Analizując powyższe wykresy można zauważyć, że faza grupowa (oraz pucharowa) na przestrzeni niespełna stulecia jest zdominowana przez drużyny europejskie. Jedynym mundialem, na którym to nie Europa była dominująca jeżeli chodzi o zwycięstwa w fazie grupowej, są mistrzostwa świata w Urugwaju, gdzie to Ameryka Południowa wiodła prym. 4 lata później we Włoszech odbyły się jedyne mistrzostwa, na których w fazie grupowej nie wygrał nikt poza Europą. Od tamtych mistrzostw rozpoczęła się hegemonia starego kontynentu i trwa do dzisiaj. Pierwszym mundialem na którym przynajmniej jeden mecz w fazie grupowej wygrały wszystkie kontynenty (oczywiście za wyjątkiem Antarktydy) jest turniej w Niemczech w 2006. Tak długo oblegane przez Amerykę Południową 2. miejsce w tej statystyce zostało przejęte przez Afrykę w zeszłorocznym mundialu w Katarze, gdzie po raz pierwszy drużyna z Afryki (Maroko) zagrała w półfinale. Różnorodność drużyn, a więc i kontynentów przez nie reprezentowanych, jest tym wyższa im bardziej rozprzestrzenia się kulturę gry w piłkę nożną. W przypadku kilku pierwszych turniejów istotnym czynnikiem była dostępność środków transportu pomiędzy kontynentami. Popularność tego sportu, problemy w komunikacji oraz pieniądze, które wciąż rosną w tej dyscyplinie, są przyczyną dominowania kilku pierwszych mundiali przez określone kontynenty – z czasem zaczęło to ulegać zmianie.

Ostatnie pytanie dotyczyło liczby remisów drużyn piłkarskich na przestrzeni lat w mistrzostwach świata. Rozkład tej zależności został przedstawiony na poniższym wykresie (Rys. 9.). Porównane zostały jedynie mecze grupowe, ponieważ jedynie w tej fazie mogły one być zakończone wynikiem remisowym. Przypuszczenia odnośnie drużyny Włoch były nietrafne, ponieważ polegały na stylu gry z kilku ostatnich mundiali. Największą liczbę remisów względem wszystkich rozegranych meczy posiada Irlandia (66,67 %) - z łącznie 9 rozegranych meczów grupowych zremisowali 6. Podobną zależność procentową można przypisać Angoli i Izraelowi, jednak te drużyny rozegrały po 3 mecze (grały na tylko jednym mundialu). W rankingu TOP 10 drużyn pod względem liczby remisów, najmniejszym względnym procentem może pochwalić się Argentyna (16,67 % - jeden mecz na sześć skończył się remisem) oraz Brazylia (18,31 %). Obie te drużyny osiągały bardzo dobre wyniki w historii mistrzostw świata - Brazylia wygrała ten turniej pięciokrotnie i jest jedynym liderem, natomiast Argentyna wygrała cztery razy finał i zajmuje drugie miejsce wraz z Niemcami. Styl grania Ameryki Południowej nie należy do defensywnych, stąd też wyniki remisowe przydarzały się stosunkowo rzadko.



Rys. 9. Liczba meczy grupowych, remisów i ich względny procentowy rozkład

Odpowiedzi na pytania:

- 1. Ile rund było w poszczególnych turniejach mistrzostw świata oraz jak na przestrzeni lat zmieniała się średnia liczba goli na mecz?**

Hipoteza była zgodna z prawdą – im większa liczba meczy w turnieju tym mniejsza średnia liczba bramek na mecz. Trzy rzeczy, które opisałem wyżej, a mogą mieć znaczenie: „większa intensywność meczów co wymusza na zawodnikach lepsze rozłożenie sił na cały turniej; poziom sędziowania oraz zasady, które na przestrzeni lat uległy zmianie; zaawansowanie piłkarzy – piłka nożna stała się bardziej popularna przez co jej poziom stale rośnie, próg wejścia jest wyższy a poziom (między pozycjami i drużynami) bardziej wyrównany”.

- 2. Czy to, że dane państwo jest organizatorem imprezy i ma po swojej stronie kibiców wpływa na wynik w danych mistrzostwach?**

Okazało się, że ma bardzo znaczący. Tylko dwa państwa (Brazylia i Hiszpania) ze wszystkich organizatorów mistrzostw świata grały lepiej na „nie swoim terenie”.

- 3. Które kontynenty dominują, jeśli chodzi o wygrane mecze oraz o całe mistrzostwa?**

Tutaj również przypuszczenia były słuszne. Wyłącznie w jednym roku – na pierwszych mistrzostwach świata w Urugwaju w 1930 roku – Europa wygrała mniej meczy grupowych niż jakikolwiek inny kontynent (Ameryka Południowa). W pozostałych była bezsprzecznym faworytem. Ostatnie mistrzostwa są również zmianą – ponieważ po 92 latach utrzymywania pierwszego lub drugiego miejsca w tej klasyfikacji Ameryka Południowa musiała ustąpić drugie miejsce Afryce, przez co sama zanotowała trzecie.

- 4. Które państwa grają najbardziej zachowawczo i miały najwięcej remisowych wyników w fazie grupowej?**

Odpowiedź w przypadku tej tezy zupełnie odbiegała od rzeczywistości. Zespołem który grał na przynajmniej 3 mistrzostwach świata i ma największy stosunek procentowy liczby zremisowanych do wygranych meczy jest Irlandia.