

ONE WORLD, ONE CUP

PREPARED BY

ARTHUR O'LEARY, CEM YILMAZ, DESMOND RYAN, MICHAEL O' LEARY

DATA VISUALISATION PROJECT BRIEF

COURSE CODE:	HDSDAJAN19A
COURSE TITLE:	HIGHER DIPLOMA IN SCIENCE IN DATA ANALYTICS
MODULE TITLE:	DATA VISUALISATION
LECTURER:	SEÁN CARROLL
ASSESSMENT:	GROUP PROJECT (3-4)
VALUE:	60%
DUE DATE:	JULY 10TH

The 1934 tournament, which took place in Italy, had the first qualifying rounds, with twenty seven nations competing for sixteen places allocated across four regions. Europe were given twelve spots, with three for the Americas and one for either Africa or Asia, a position that was ably filled by Egypt. The 1938 tournament in France had a similar composition (with the Dutch East Indies being thrashed 6-0 in their sole World Cup performance), but the worldwide competition was put on hiatus during the Second World War, despite Germany's lone bid to host a 1942 World Cup. In Brazil in 1950, the FIFA world cup continued its long history of alternating between Europe and South America, with the winning team almost always coming from the host continent, often the host team.

Over the next seven decades, the game has spread rapidly across the entire world, and the world cup finals have expanded to accommodate the growth. In 1982, twenty four teams contested the trophy and in 1998 it grew to thirty two. While some have criticised the bloating of the tournament and the consequent reduction in the number of quality matches, there can be no denying that the World Cup aims to meet the mankind's urge to enact a kind of pantomime world war. In total, 81 team have played in at least one World Cup final and an extraordinary 211 countries will attempt to qualify for the 2022 World Cup, meaning that there is scarcely an outpost left in the world where football is not played as an international sport. With demand now at capacity the powers-that-be have decided to allow 48 teams into the 2026 World Cup, in an attempt to grow the game further and to share the boundless wealth and prestige associated with the world's one truly global sport.

DESCRIPTION OF THE DATA



The analysed data was extracted from a dataset containing details of all 40451 international football matches played between 1872 and June 2019.[i] Details were included for the date, location, home team, away team, scoreline and tournament, which was included as a factor with 104 levels.

Thus it was possible to extract all 900 World Cup matches (1930-2018) and the associated 7100 World Cup Qualifying matches (1933-2017).

Data analysis was initially performed in R Studio and, while the dataset was rather clean, it was necessary to adjust the date for each game, using the lubridate library, as part of the tidyverse package.

The year could be extracted from the original DD/MM/YYYY format and this was binded to the main data frame as a separate column, then turned into a factor. It was thus possible to subdivide the data into 21 world cup tournaments and analyse the games as appropriate. Several kinds of analysis were performed, primarily based around the performance of each team at the world cup and the number of goals scored.

Teams contesting

World cup matches are neutral unless the host is playing so the dataframe's column for home and away teams were more of a nuisance than an aid. In order to isolate the participants it was necessary to search the Home and Away columns for unique names and then list them separately. A function was written to do this and it was run through the data frame associated with each of the 21 World Cups. Lists of participants then had to be checked against other data sets (Wikipedia was used as it is particularly reliable on matters as incontrovertible as World Cup results) This process involved the use of Google Sheets to extract data from tables embedded in HTML webpages.[i] No conflicts were found throughout the project suggesting that the dataset in use is particularly reliable.



Goals Scored

With the dataframe split into 21 subsets it was easy to analyse out the number of goals scored in each tournament, averages, ranges, etc. However, the fact that the winners and losers were not specified in the meant that the data had to be rearranged in order count the number of 0-0 draws, 1-0 wins, and so on for every scoreline up to 7-5.

Problems with the data and the software.

One of the problems with presenting the data has been the ever changing nature of the world's nation states. Many countries have come and gone since 1930 and several have merged, split or changed name. The imported dataset had already retroactively fixed some, but far from all, of these names. The easiest cases involve simple name changes: DR Congo played in the 1974 world cup under the name of Zaïre, and that has simply been swapped out. Similarly, the erstwhile Irish Free State (who failed to qualify in 1934 or 1938) became known as Republic of Ireland during qualification for the 1950 World Cup (to which they refused a late invitation). Things get more complex when a country splits — Czechoslovakia, who lost two World Cup finals (1934 and 1962) — has become Czech Republic and Slovakia. More complex again has been the partitioning of Yugoslavia who performed excellently for sixty years.

They have been replaced by Croatia (even more successful), Serbia (under a few guises of their own), Slovenia, Bosnia & Herzegovina among others who have never qualified (e.g. North Macedonia). And of course the Soviet Union devolved into more than a dozen football teams, of whom only Russia and Ukraine have qualified for in the world cup. Significantly, the dataset uses the name Russia for both present-day Russia and the Soviet Union.

3.5 Auxiliary and irregular data.

Along with data validation, it was occasionally necessary to use information taken from other sources. For example, the dataset did not indicate whether a game took place in Europe (governed by UEFA) or South America (CONMEBOL), etc. Again, Google Sheets was used to import HTML-embedded data on the topic.

Another approach would have been to tag the original dataframe with a factor column representing either the continent or, more appropriately, the governing body to which each country is attached (e.g. Turkey and Kazakhstan are governed by UEFA, while Surinam is part of CONCACAF, not CONMEBOL).

Frequently, our project was met with a dilemma: should we learn more R code in order to work with the data, thus taking unknown amounts of learning time, or should we use familiar techniques in Excel for which timeframes could be established. Different members took different approaches allowing us to aim for the overall best of both world. In some instances, there were some unforeseen problems in Excel which resulted in some interminably slow and cumbersome delays in organising data (e.g. the Bar Chart Race, below).

4. Description of data analysis



4.1 Auxiliary and irregular data.

We calculated that the average goal count per tournament since 1930 is 3.08. Recent tournaments (post 1970) have shown that the goals per tournament have dramatically declined from this figure, we can only postulate that this is due to the overall number of teams in the competition increasing or a change in the style of game play, to a more defensive style.

4.2 Goals per game

Goals per game – There were two fields available in the data set for home and away goals. It was relatively straightforward to merge these two fields to get a final result for each game. Seeing as home and away goals are largely irrelevant in a World Cup game, we were able to go back and add up results that ended in 1-2 and 2-1 to a home team, for example. This was the most manual stage of the process. Once this had been completed we were able to use a pivot table to see the number of instances of a particular scoreline that taken place across all World Cups.

4.3.1 Performance Aggregate

A script in Microsoft Excel was used to go through the results at speed and deduce the outcome of each game.

=if(HomeGoal > AwayGoals, "Yes", "No")

The next step was to go through the list of games and decide what stage of the world cup the game had taken place in. The date field in the dataset was a big help here, with the most important games at the end of the tournament, and different rounds occurring within a cluster of dates in each tournament.

A ranking system was then decided, where a winning result at different stages of the tournament would be worth a progressively greater value:

Once every team in every game had been given a value based on their performance at a particular stage of the tournament, we could stack all this data and use a pivot table to aggregate the performance of each team across all World Cups.



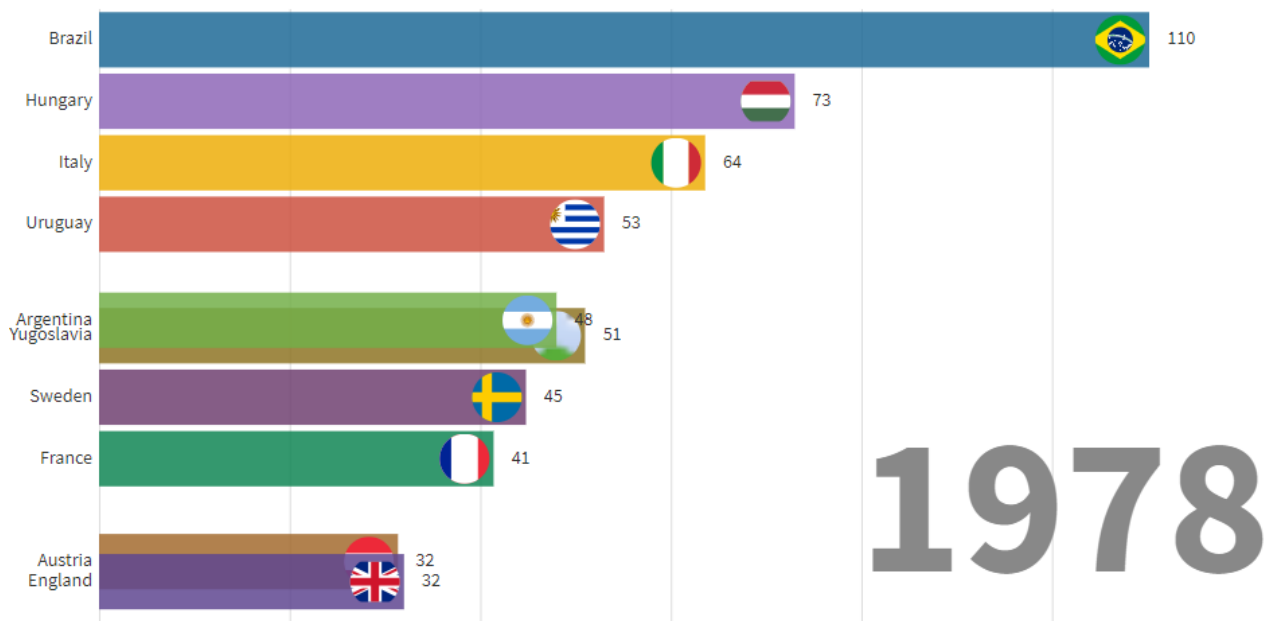
4.3.2 Data Issues

World Cup, the biggest challenge was looking back at the old games and deciding whether a game was a quarter final or round of 16, and that the corresponding points value was being applied properly.

A second issue with the dataset was that the results only showed the final scoreline, so that when games were decided by penalties the winner was not made explicit. We got around this issue by giving both teams points for a draw, while the winning team would have the opportunity to gain more points in the next round. This only left the two finals that were decided by penalties (1994 and 2006) to look up, so that the correct teams were given points for winning the World Cup.



4.3.3 Bar Chart Race – Aggregated Performance across all world cups



There were several challenges faced in producing the bar race animation. Flourish.studio requires CSV files to be formatted in a very specific way. This required hours of data wrangling using Excel to prepare a CSV file in its required format. Due to using different sources of data to extract the information required for this bar race, we failed to add the 2018 world cup goal numbers to the animation. Therefore, the number of goals presented is in fact not the final goal count of countries and a final chart would require one last act of data merging. Joining different data sets for the specific data was too challenging to complete given the tight time frame to complete this task, and again showed up the problem of using combined approach (Excel and R Studio). Preparing the CSV file required hours of completing online excel YouTube tutorials. This method of remedy was applied to all excel related challenges faced during the preparation of the CSV files. Link here. <https://public.flourish.studio/visualisation/481831/>

5. Description of the visual design process

5.1 Average goals per game: several approaches

Armed with the data, it was easy to calculate the average number goals per game in each tournament. The issue was how best to present this information in a way that was informative (easy with a barplot) and interesting (not so easy). The first thing was to do a basic barplot, showing the average goals per game. The next task was to work out how to demonstrate subpatterns in a graph where there are two very distinct groups (pre-1962, shown in the first six bars, and 1962-2018, the rest).

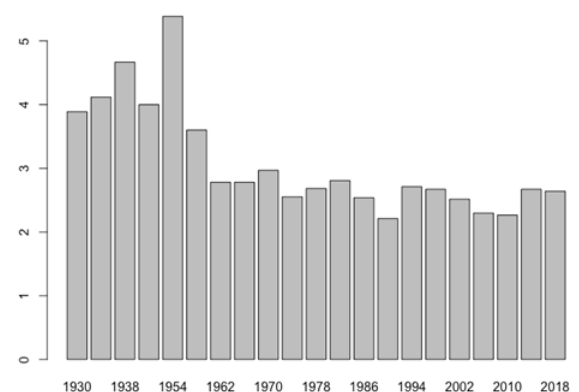


Figure 2: Bar chart showing average goals per game in each of the 21 World Cup finals

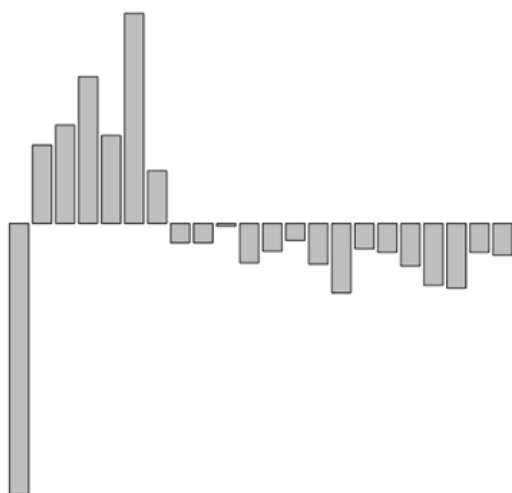
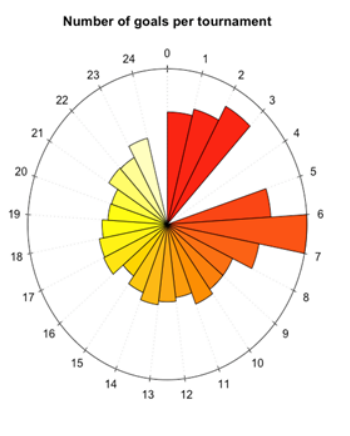


Figure 3: Bar chart showing the deviation from the average number of goals (3.08) by tournament

One combined graph, designed using ggplot in R, was to include bars showing some averages: the average across all tournaments (shown in thick yellow), the average between 1962 and 2018 (a normal-looking distribution as noted earlier). The advantage here is that we could use bars to show two distinct averages — the overall average and the 1962-2018 average. This allows us to compare like with like, and separate the modern world cups from the wilder earlier editions.

Such options are hardlyt innovative so the next iteration was to try some different methods of showing this kind of data. One option was the clock plot shown below.



A neat way of showing the contrast was to show the distances from the mean, above and below. In barplot 2, the leftmost column shows the average number of goals per game across the 21 tournaments (3.08), and the two subgroups are neatly shown as being unequivocally above or below it. The problem with this plot is that it is hard to interpret for the reader (and the horizontal plot was even harder to comprehend).

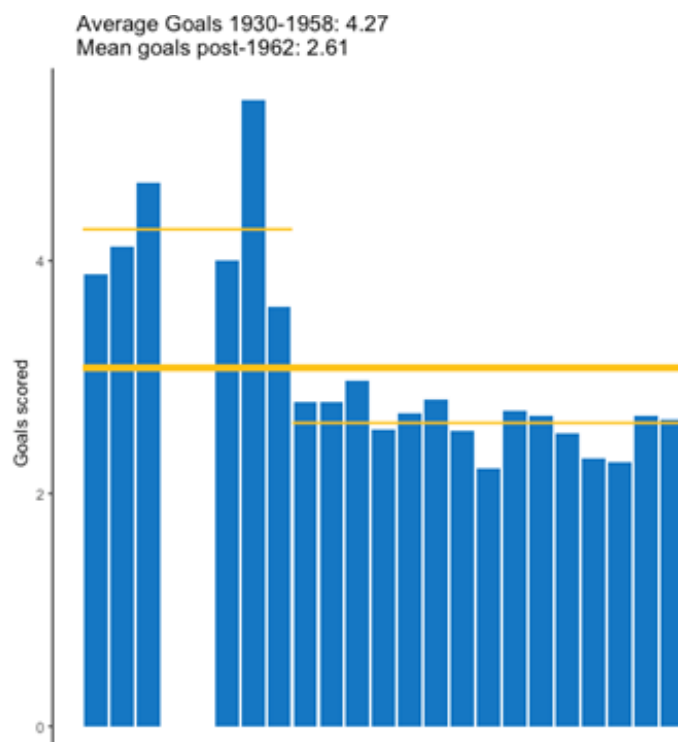


Fig 4. Bar chart showing average goals per game in each of the 21 World Cup finals, including overall and local means.

In this unusual plot, the edition of the world cup is shown by the hour number (the code[i] did not allow for the alteration of the numbers). The heat map marks the passing of time and this colour scheme was chosen because the general tendency was from many goals (red) to few (yellow). The chart is useful for showing several distinct phases of time: the three pre-way World Cups, the wartime hiatus, the onslaught of goals in the 1950s, and the relative steadiness of the last fifteen editions of the World Cup.

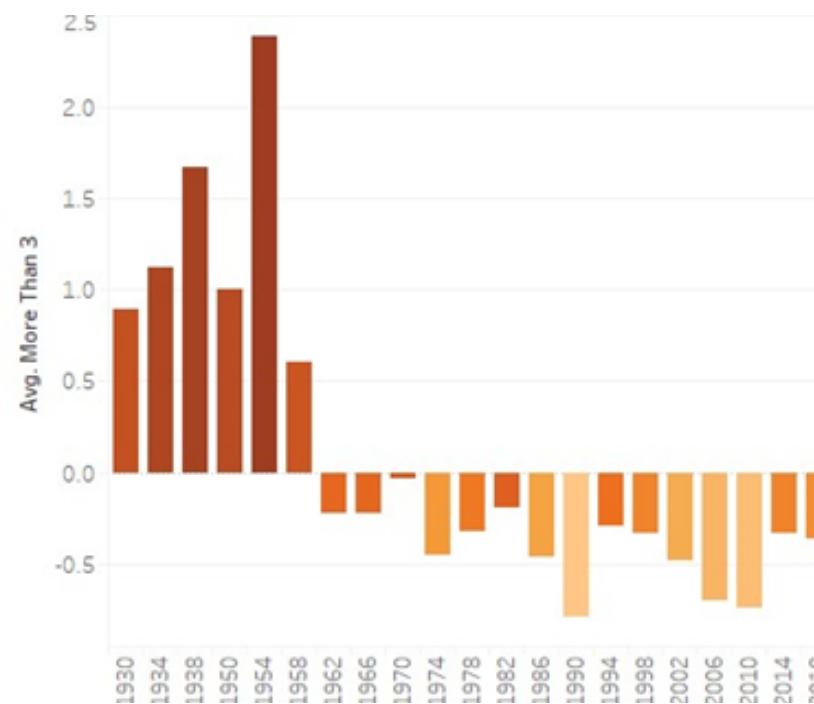


Figure 6: Bar chart showing the deviation from the average number of goals (3.08) by tournament. The heated colour scheme shows the number of goals

However, the plot was not easy to understand for our interactive users who needed quite a bit of background explanation. Nor is it good for showing subtle differences across distant tournaments (e.g. were there fewer more goals scored in 1990 or 2006 (i.e. edition 16 v 20)?

In the end, we decided to use a combination of all options in the interactive visualization. We used a coloured version of Barplot 2, and users can hover over each bar in order to its details in print, solving the problem of having a 'floating' bar with no zero point. The clock map proved to be useful in demonstrating the usefulness of a heatmap colour scheme, reflecting the change from many to few goals. The plan was also to use this graph in the infographic but it was cut due to space constraints.

5.2 Bubble plots: ideal for showing changing scorelines over time.

A bubble plot shows three pieces of information and this is ideal for showing scorelines: each team's score and the frequency of occurrence. In order to show changes over time, we used two bubble plots. One for the first six world cups and another for the rest.

Shape

Efforts were first made to use Tableau to produce the bubble plots of scoreline frequencies in the World Cup. However, it was quickly established that R Studio would provide more options that would be easier to control. The ggplot library was used and the data was prepared for the purpose. The easy part was plotting the information but the design required many amendments. Firstly, it was necessary to distinguish between draws and wins. ggplot provides an option for putting lines onto graphs but it was felt that an arrow would convey the

information better due to the possibility of higher-score draws (e.g. 5-5 or even 6-6).

The arrow also draws the eye to some important outliers among the wins (e.g. 6-5, 7-5).

Colours

The choice of yellow line follows the project's palette, as does the white background (which required the changing of the default ggplot background). Draws are written in yellow, like their line, but black was kept for the word Wins in order to connect the words meaning to the unhighlighted black dots.

Size

The area of the bubbles demonstrates the proportions except that a multiplicative function was added to increase the size of the outliers (which might otherwise be invisible).

Importantly, the two bubble plots are based on differently sized samples but the proportional nature of the plot resolves his problem and saved us the need to use percentages, reducing the amount of information stored. The problem is that this requires the reader to be able to judge frequency by circular area, something that can be difficult for many people.



Position / Orientation

The plot may well work better if it were reflected 45°. We asked several people and some preferred the chart with the winning team on the y-axis. The tension stems from two familiar intuitions. Football scores are usually printed and read from left to right (although away wins may not be). By contrast, cartesian co-ordinates, upon which ggplot is based, are inputted with the x-axis first and y-axis second. Both directions have a problem. Putting winners on the y-axis leads the eye towards the top left (rare 10-0 wins, dropping down to 0-0 draws) whereas putting them on the x-axis leads the reader from the (relatively common) 0-0 draws past the commonplace 1-0 and 2-0 wins and then to the rarer positions.

Either way, this plot has the problem of requiring extensive explanation as most people are not used to vector plots shown through three dimensions (x, y and size).

5.3 Bar Race

Why chose a bar race?

The bar race provided by flourish.studio illustrated the number of goals scored by each team throughout the fifa world cup year by year. The decision to create a bar race illustrating the number of goals scored by each team in this manner was to allow the observer to intuitively understand which team was in the lead in the number of goals scored, thus providing a proxy measure of success. By combining the element of time with the total number of goals scored by each time, the bar race objectively presents how each country team compares with other teams in terms of number of goals scored adding a competitive element to the animation.

Speed

The ability to change the speed of the animation improves eligibility. There were in total 19 world cup matches meaning 19 different instances. The duration of the animation with only 19 instances can end quite quickly, diminishing the intelligibility of the bar race. By reducing the speed, the intelligibility of the animation is improved.

Colors

Each country has its own unique color which remains consistent throughout the whole animation. This was specifically set to determine consistency during the animation. The colors themselves selected are bright and colorful colours to give a serious but also playful theming to the animation.

Size

Size was definitely an issue that is omnipresent in all bar race charts that include a large number of variables, or in this case, number of country teams. In order to make the animation aesthetically appealing, intuitive and tidy only 10 teams could be presented in any given time. This of course restricts the number of teams that can be presented in any given time.

6 Discussion

6.1 The growth of world football.

81 states have been represented at the football world cup, as shown in the heatmap in Figure 1, although this of course includes some overlap discussed in Section 3.3. Football has spread enormously in the last fifty years and this has surely been helped by the expansion of the tournament in 1982 to include twenty four teams.

That year, allocations were doubled for Africa, Asia and North America. In France in 1998, the tournament swelled to 32 teams with 13-14 slots being available for these regions (alongside Oceania / OFC). The demand for places is enormous and fully 210 countries attempted to qualify for thirty two place in the 2018 World Cup. Accordingly, there will be room for 48 teams at the 2026 World Cup, with nine from Africa and eight from Asia, ensuring that these continents combined will outnumber European representation

The tournament will be hosted by USA, Canada and, for a record third time, Mexico. While the two English-speaking nations may not be traditional footballing powerhouses (USA's bronze medal in 1930, notwithstanding), their combined market is huge and women's football is particularly popular and successful. The US have thrice been World Cup winners in a row, fact that will impact on the overall status of the sport in what is becoming one of football's largest markets. While many have criticised this inflation [MOU1], arguing that the quality of the finals matches will be affected, it is hard to deny that demand is growing for a greater supply of final places.

	Allocated places		No countries attempting qualification for 2022
	2022	2026	
Europe (UEFA)	13 + hosts	16	55
Africa (CAF)	5	9	54
Asia (AFC)	4	8	46
South America (CONMEBOL)	4	6	10
North America &/Caribbean (CONCACAF)	3	6 inc. three hosts)	35
Oceania (OFC)	0	1	11
Others via play-off	2	2	NA
Total	32	48	211

The worldwide spread of the game is best summarised by the allocation of places for the 2022 and 2026 tournament. It remains to be seen whether these continents will find success to match their numerical growth, but the history of international football has shown a gradual levelling of the playing field among the chasing pack (see the infographic on aggregated performance). No African team has yet been to the Semi-Finals and South Korea, on home soil in 2002, were the only Asian team to get that far (Turkey being a UEFA nation).

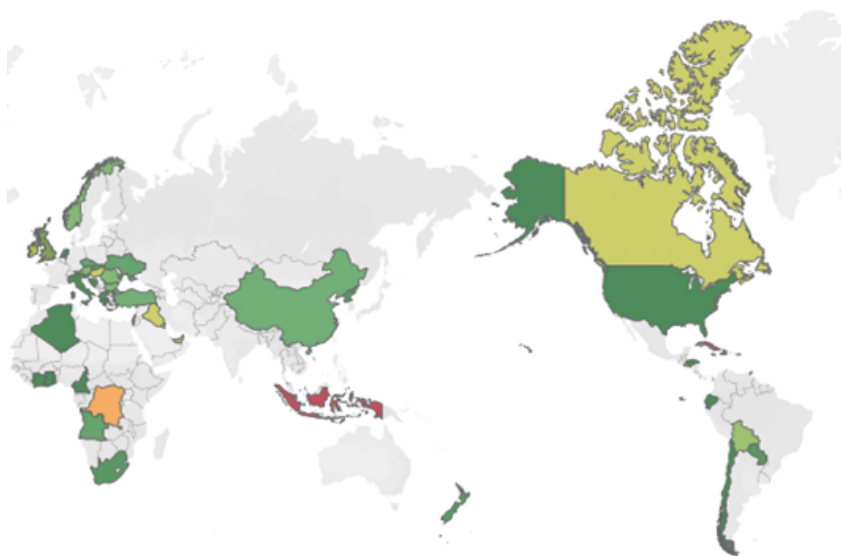


Figure 7: Map showing how long it has been since various countries have played in the world cup. Red is more, green is less.

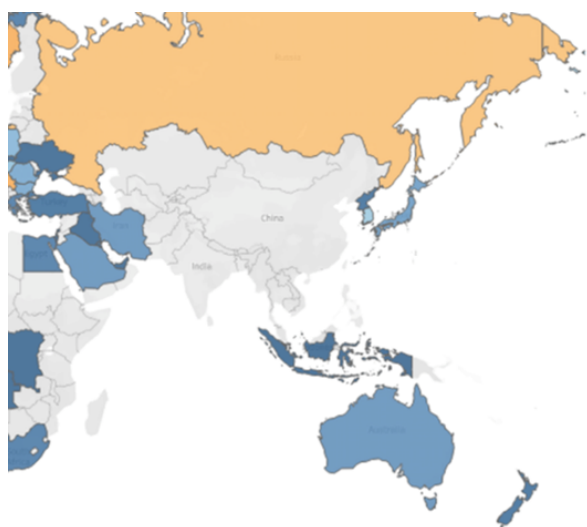


Figure 8: Map showing the paucity of appearances among nations around the Indian Ocean

Finally, the growth of football does not mean that every time has been steadily improving. The ‘Qualification Drought’ graphic shows us, despite, a large growth in confederation participation there are many teams who have struggled to capitalize on early success. Indonesia (shown in dark red), were represented in 1938 by The Dutch East Indies, and they have not qualified since their solitary loss that year. And even Scotland, once a major force in football, have not qualified since 1990. Thus we see that many nations struggle to keep pace with the dynamics of the subtly changing sport along with the administrative issues that are required for consistent success.

It should also be noted that large swathes of the world’s population have been severely under-represented at the world cup. The entirety of South Asia, stretching from Pakistan to China and down to Indonesia (a region in which half of the world’s population live) have scarcely been represented at any finals. A close reading of the interactive chart will demonstrate this in detail but the picture below shows a close-up of this region.

Many former Soviet states have struggled to qualify and nearby Finland have never qualified, despite trying since 1934. East Africa has not been successful at all and this highlights its historic ties with countries of the Indian Ocean, unlike West Africa where teams have been relatively successful (Morocco, Cameroon, Nigeria, Senegal), perhaps due to closer connections with Europe and South America.

The entire Indian subcontinent is a football desert, as is South East Asia. In Oceania, association football must compete against other footballing codes (e.g. rugby union, rugby league, Australian rules) and this is true of most English-speaking countries: USA, Canada, Ireland and Wales all play football codes which originated in English public schools. This fact also accounts for the Indian subcontinent and the West Indies which are cricketing strongholds.

It remains to be seen whether football takes hold in China, South-East Asia and the Middle East. Support for the European club teams is enormous in these countries but perhaps that sedentary interest will not be manifested as success in international football.

6.1.2 Host nations and the spread of power (or lack thereof).

Despite the spread of teams playing football, only eight nations have won the competition (assuming we take Germany and West Germany as one), and power remains concentrated among South American teams and, even more than ever, European teams. Brazil have won the trophy five times, although not since 2002, while Italy and Germany have four, wins apiece. Uruguay, Argentina and France have two each, while England and Spain have both won the tournament once. Perennial bridesmaids Netherlands and Argentina have lost three finals apiece.

Home advantage has been very helpful with six of the above teams winning the tournament on home soil (all bar Spain and Brazil) but perhaps even more strikingly, the tournament was won by a home continent almost every year until the tournament began being hosted around the world (beginning with Japan and South Korea in 2002, where the latter team made the semi-finals, a first for an Asian team). But the current trend is that a European team will now win it, no matter where it is held (France won in Russia in 2018, Germany won in Brazil in 2014, Spain won in South Africa in 2010, etc).

6.2 Goals, goals, goals

The most glaring fact about goals in the World Cup is that there used to be a lot more of them. Until 1958, there were more than four goals per game on average, but since 1962, there has only been 2.68 goals per game, with the 1990 world cup averaging a measly 2.28. The bar chart above shows the average number of goals per game in each World Cup. The thick yellow bar shows the overall average of 3.08 goals per game, while the thin yellow bar shows the average since 1962.

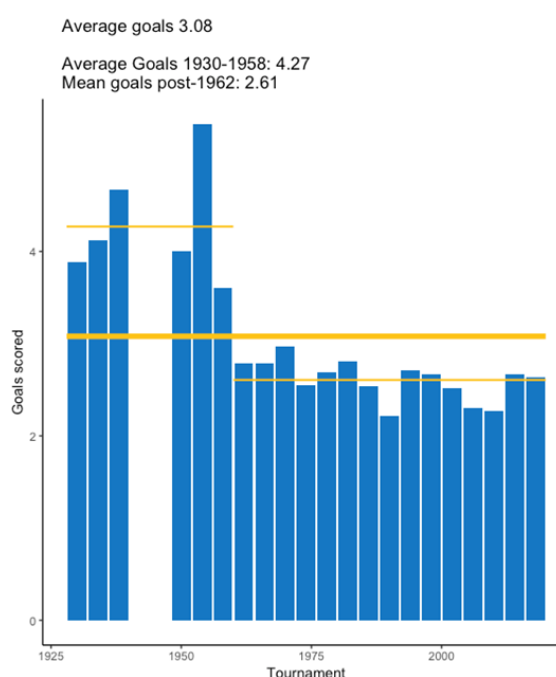


Figure 9: Bar chart showing the deviation from the average number of goals (3.08) by tournament. The heated colour scheme shows the number of goals.

1954 was a bonanza year (which featured a very prolific Hungarian team) but the reduction began suddenly in 1962 and goalscoring has been relatively steady since then. The tempting hypothesis is to say that rule changes after 1990 resulted in major changes in the sport. For example, goalkeepers were no longer allowed to pick up a back pass and the offside law [MOU1] was altered to allow players in line with the defender to be considered onside. The slight surge in goals in 1994 and 1998 would suggest that the rulemakers were vindicated. And yet, a Shapiro test for normality suggests that the sample of average goals per tournament between 1962 and 2018 is normally distributed ($p = 0.33$, $N = 15$). The null hypothesis — that nothing significant has happened in the last 15 world cups — cannot be rejected. The lack of goals in 1990 may simply be down to chance. There may also be other factors

The question remains as to why defences have reigned supreme since 1962. Jonathon Wilson (2014), in his history of football tactics, notes the emergence of many new defensive formations in the mid-1960s (e.g. Inter Milan's catenaccio) but this does not explain the sudden drop off in goals scored in this period.

Home advantage has been very helpful with six of the above teams winning the tournament on home soil (all bar Spain and Brazil) but perhaps even more strikingly, the tournament was won by a home continent almost every year until the tournament began being hosted around the world (beginning with Japan and South Korea in 2002, where the latter team made the semi-finals, a first for an Asian team). But the current trend is that a European team will now win it, no matter where it is held (France won in Russia in 2018, Germany won in Brazil in 2014, Spain won in South Africa in 2010, etc).

6.3 The changing scorelines in football

Perhaps more revealing are the actual scorelines in the games. The bubble plot belows show the number of games with each scoreline, with 0-0 draws (9%) being at the origin. However, it only shows the games between 1962 and 2018 after the average number of goals had dropped dramatically.

The x-axis shows the number of goals scored by the winning team (between 1 and 10 goals) while the y-axis shows the goals scored by the losing team. Draws are on a 45° angle where $x = y$, and the distribution of the data appears to be rather even. The most common result is 1-0 and the vast majority of games have no more than two goals apiece. Games with more than six goals are very rare (3.3%).

Outliers do exist of course, but not so many in recent decades. 1962 saw a 4-4 draw between Colombia and Soviet Union in 1962, and there have also been two nine-goal slaughterings. Hungary's 10-1 drubbing of El Salvador in 1982 and, in 1974, Yugoslavia's 9-0 trouncing of Zaïre. There have only been two extreme outliers this century: Germany's 7-1 demolition of Brazil in the 2014 semi-final and their 8-0 win over Saudi Arabia in 2002.

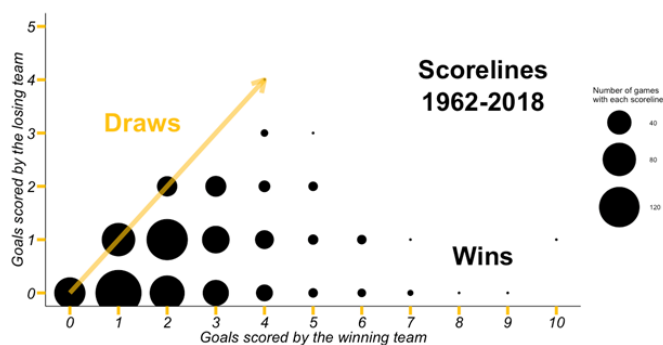


Figure 10: Bubble chart showing the frequency of every scoreline between 1962 and 2018.

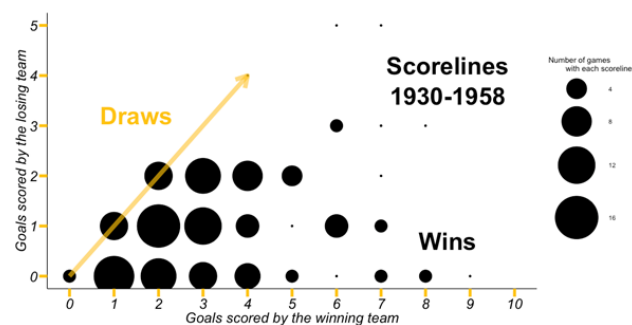


Figure 11: Bubble chart showing the frequency of every scoreline between 1930 and 1958.

The first six world cups show a very different picture. There were very few 0-0 draws (1.7%) and the most common result was 2-1 (12.9%). Games with more than seven goals are relatively common (12.7%) as are the expected outliers, including Brazil's 6-5 win against Poland in 1938, Austria's 7-5 defeat of Switzerland in 1954 (the highest number of goals in a world cup game) and even Brazil's 5-2 win against Sweden in the 1958 World Cup final.

It is impossible to know exactly why the number of goals has reduced so dramatically in football but obviously defences have improved. Other things have changed however, including the organisation of teams (by managerial teams) and even the equipment (e.g. the ball itself, boots, goalkeepers' gloves, etc.).

One might also claim that there are fewer weak teams in the world cup, resulting in many thrashings. Yet the bubble plot of the early years shows many games where the losing team scored two or more goals. Thus, equality is not solely to blame.

Further study would be necessary to unravel the myriad of factors contributing to the nature of the game.

More goals?

More teams means more games which means more goals. Independent of that, it does appear that there will be more goals in future games, regardless of the number of teams. The time-series plot below (created in Microsoft Excel) shows the projected number of goals scored in future world cups. The extra number of teams competing accounts for the general upward trend and a fuller analysis would require us to weight editions since 1962 more heavily.



Figure 12 Time series analysis showing a first draft of projected goals in future World Cups.

Conclusion

It is clear that the world cup has grown in size in many ways. There are more teams representing more countries and thus more matches. Yet there is still just one winner, one finalist and one bronze medalist. Eight teams have dominated the tournament but our performance aggregate shows some deeper trends that might not otherwise be visible. The Netherlands have never won it but it may be argued that they have performed better than two-time winners Uruguay. Czechoslovakia have played two finals, and many other teams have gone that far, including Hungary, Sweden and Croatia, who have improved on Yugoslavia's excellent results. The semi-finalists have cast a wide net with South Korea representing Asia (CAF) in 2002 and Turkey (UEFA) in the same year. North Korea made the 1966 quarter final (where they led Portugal by three goals) and were the first non-European, non-American team to progress past the group stages. African teams have grown steadily, with Cameroon beating then-champions Argentina on their way to the 1990 quarter-final, an African first, and they were emulated by Ghana in 2010 (in Africa, it should be noted). Some teams have been whipping boys and several teams have scored zero goals in their solitary world cup performances: UAE, Togo, Panama, Iraq and Indonesia.

The final question to ask is whether the globalisation of the game will eventually be reflected in results for countries beyond the traditional powerhouses. The overall projection looks good but very, very slow. However, the importance of hosting a tournament should not be overlooked, nor even having a tournament near your home. Thus the spread of host nations into Africa, Asia and North America will undoubtedly aid their development. Further study would be necessary to unravel the myriad of factors contributing to the nature of the game.

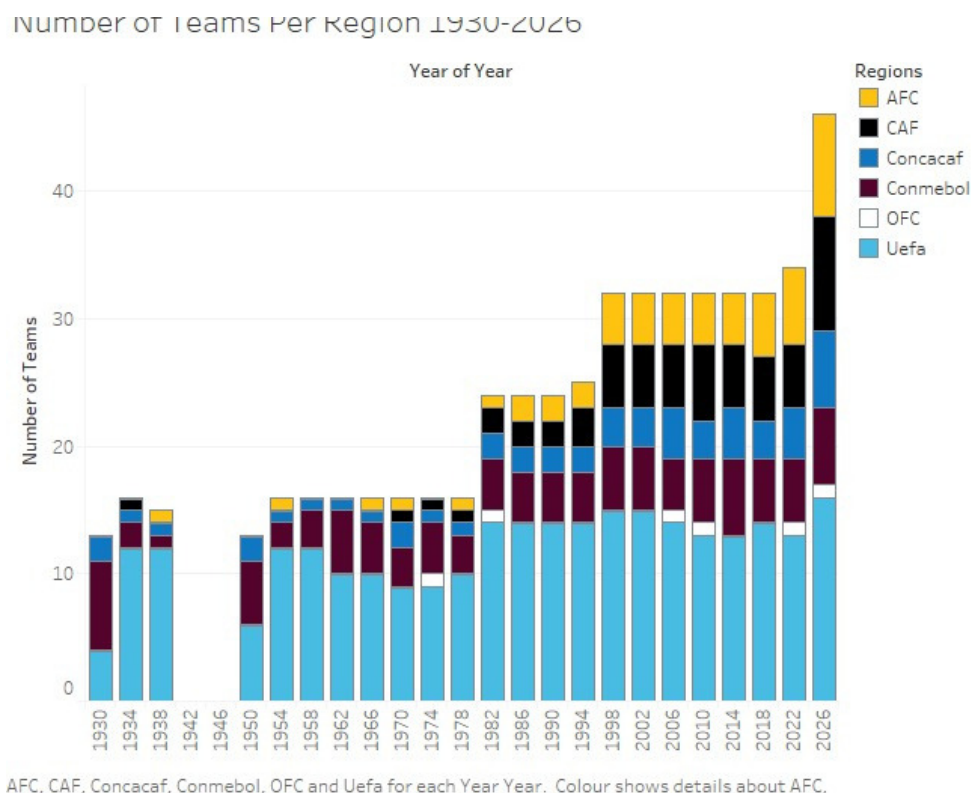


Figure 13. Stacked barplot showing the growth in the number of teams representing each region. Note the growth of Asia (yellow) and Africa (black) relative to Europe (light blue) and South America (maroon). North America / Caribbean has also grown but Oceania scarcely features.

References

Corbett J. and Crouch. T (2014). The world cup: a complete history. 3rd Edn. Aurum Press: London.

Orbis (1990). World cup '90: a sticker collection.

Uli & Hesse (2013). Tor: the history of German football.

Wilson, J. (2014). Inverting the pyramid: a history of football tactics.

[1] The entire dataset is available here <https://www.kaggle.com/martj42/international-football-results-from-1872-to-2017>, last checked on 10 July, 2019.

[1] For example of Google sheets code:

```
=IMPORTHTML("https://en.wikipedia.org/wiki/National_team_appearances_in_the_FIFA_World_Cup","table", 8)
```

[1] Code adapted from: <http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html#Violin%20Plot>

