Greatness in Sports:
Understanding and Visualizing its Patterns
Article Draft, V1

Benjmamin Leinwand and Matthew Murray

Introduction

A common phrase that is emphasized among sports fans and expert calls to not take greatness for granted. While the sports themselves are (hopefully) lasting, the athletes come and go, with their careers seeming like a brief blip in time when it is all said and done. Baseball fans listen to stories of the Great Bambino from their elders, wondering if they will ever see a player or a talent burst on to the scene again. Meanwhile, basketball fans seem to engage in an everlasting debate between Michael Jordan and Lebron James.

Nonetheless, with the seemingly endless abundance of talent on display today, coupled with the advances being made in athlete performance and recovery that have athletes performing at the highest levels ever, one has to call to question: *can* we take greatness for granted? Are there more greatest of all time (GOAT) candidates than one may intuitively believe there to be? Are there patterns in which greatness occurs in one sport, and a lack thereof in another? Do sports with certain qualities (e.g: contact vs. non-contact) tend to attract more GOAT candidates than another? In this paper, we seek to provide the answers to these questions.

Data - Description and Limitations

Due to the unique nature of this project, we had to impute our own data. While time consuming, doing so allows us to be fully transparent and elucidate on all the assumptions that we inevitably had to make while creating the datasets. We decided to compile lists of men's GOAT candidates from the following sports:

"Subjective" Sports

1. Soccer
2. Cricket
3. Tennis
4. Basketball
5. Baseball
6. Football
7. Ice Hockey
8. Golf

"Objective" Sports

9. Chess

10. Swimming
11. Running

The inclusion of football, baseball, basketball, and ice hockey can be attributed to the fact that they are the 4 of the 5 most popular sports in the United States (Gallup 2017). Golf and tennis were included because professionals in both sports compete in four "major" tournaments each year. Soccer and cricket were included due to their immense popularity and following worldwide (Media 2017). Additionally, we decided to impute data for three "objective" sports - chess, swimmming, and running. Put simply, these sports have very well-accepted, objective metrics that are more-than-often referred to when discussing the best player in that sport. The inclusion of these more objectively measured sports serves as litmus test from the results of the subjective sports.

Another important point is that the data spans from 1945 to the present. One key limitation of this approach that we would like to point out is that after 2010, we are likely underestimating the number of GOATs just because most of the great players have not been playing long enough to truly be in the GOAT conversation. (A perfect example of such a player would be Patrick Mahomes of the Kansas City Chiefs. The choice of 1945 can be attributed to the fact that at this time, the MLB (1876) and NFL (1920) were already founded, while the Basketball Association of America (which eventually became the NBA) was founded shortly thereafter in 1946. Therefore, one can argue that the three most popular sports in North America (in addition to soccer) were already up and running by that time period. Additionally, 1945 marked the end of World War II and the advent of a period of economic flourishing in the Western world.

Out of our valid concern of bogging down - and slowly lulling to sleep - most readers, we have not included the full list of statistical liberties - including how we determined a player's active years, the different criteria we used for including players in our dataset, etc. - that we took while creating the dataset in this article. These can instead be found at our GitHub repository. Please refer to the file "Data Collection Assumptions.pdf."

Disclaimer - What we are NOT Trying to Do

One point that we would like to *avidly* emphasize before getting into the details of the paper is that we are *NOT* looking to answer the question of who is the GOAT in any sport. Said question - while incredibly interesting and thought provoking - has been discussed ad nauseam at every bar, barber shop, and online forum imaginable. In fact, the ubiquity of the question is a testament to the fact that it can almost never be answered objectively. We are instead looking to examine the frequency, patterns, and cadence in which legitimate GOAT candidates appear in the most prominent and popular sports. In doing so, we had to inevitably - find GOAT candidates and sought to do so by using very reputable and defensible sources and metrics.

Additionally, we are aware and cognizant of the fact that if a reader were to conduct a similar analysis, the results and data may be very different. However, our goal in this project is not to conduct an analysis that is extremely robust/resistant to sensitivity tests, but rather, one that is informative, insightful, and interesting. If 25 different researchers were to conduct this analysis, each dataset and approach may be very different, and we are okay with that fact. Like many statistical analyses, answering our questions of interest is an art, not a science.

Notable Findings

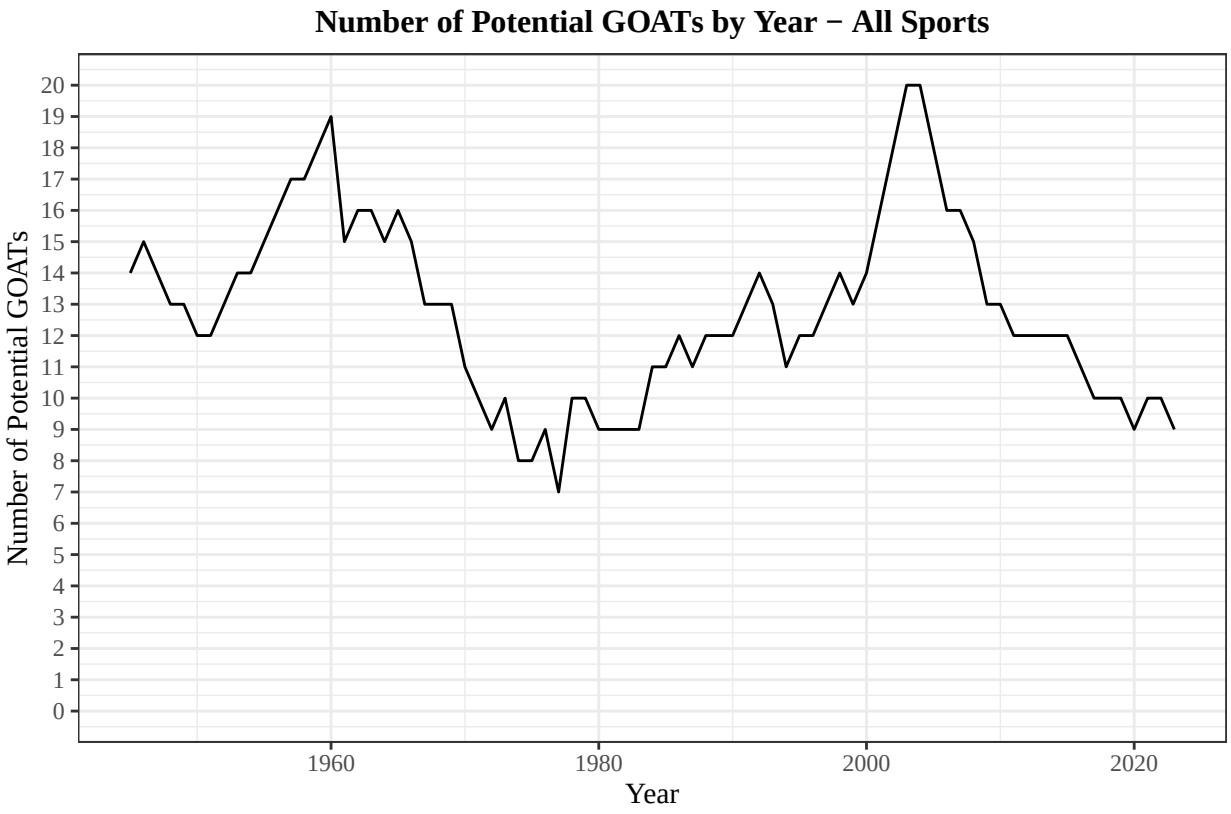

Figure 1

Key Observations:

- Min: 7 (1977)
- Max: 20 (2003-2004)
- Mean: 12.78
- Median: 13

Finding 1

There are 3 main cadences at which GOAT candidates appear. They - as well as the sports that fall into each grouping - are as follows:

1. GOAT candidates are essentially one-offs, or statistical anomalies - Baseball, Cricket

2. GOAT candidates occur sporadically without any legitimate pattern, which aligns with our prior beliefs - Golf, Ice Hockey

3. GOAT candidates either occur more frequently OR in larger quantities than we previously expected (corroborating our hypothesis that greatness can be taken for granted) - Basketball, Football, Soccer, Tennis

Finding 2

The number of years during which a sport has a GOAT candidate does NOT appear to be affected by the sport's average career length OR whether the sport is contact or not.

Career Length

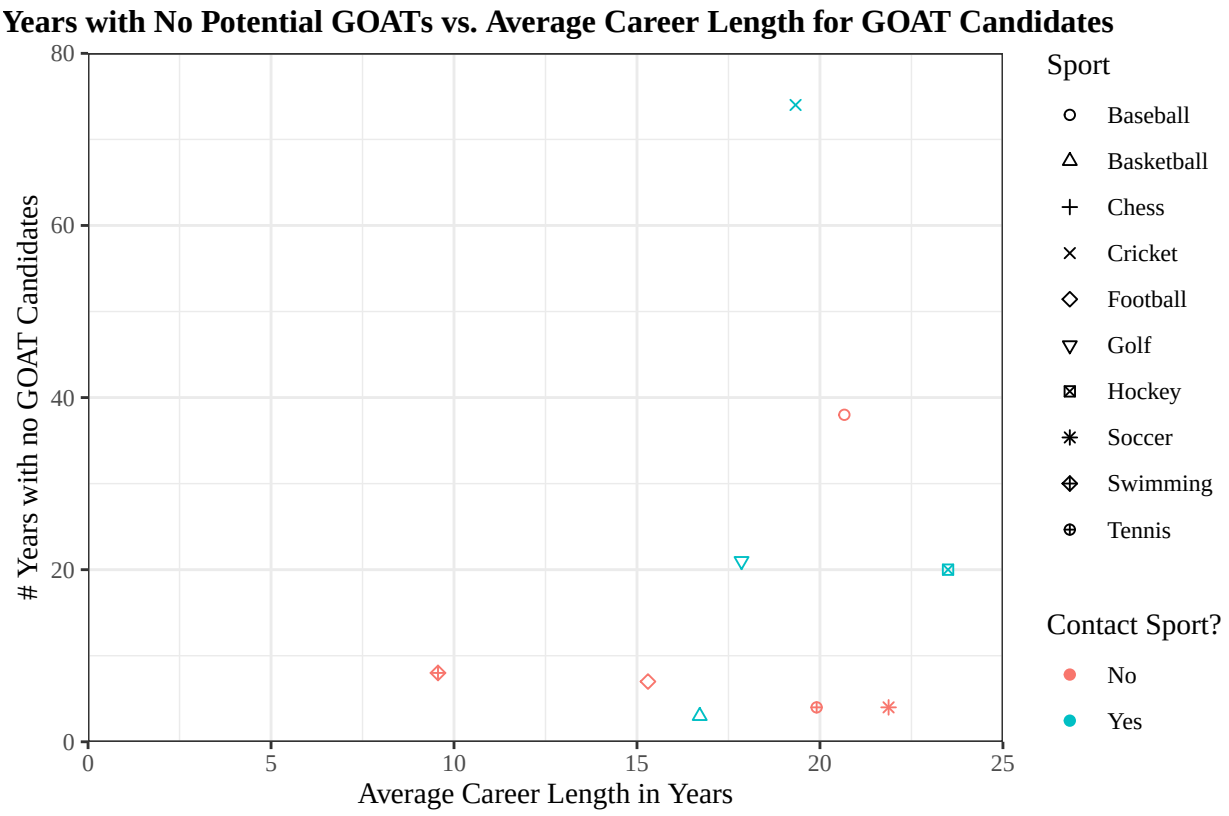

Figure 2

Based on **Figure 2**, no trends in particular seem to stick out. Although football and basketball (both of which are contact sports) both seem to have shorter average career lengths, one cannot say this for all contact sports compared to non-contact sports. However, a potential limitation/inconsistency for golf (a non-contact sport) is that the average career length may appear to be shorter than it actually is because we used the first and last years a player won a major for a career span.

Intuitively, one would expect sports with the *shortest* career spans to have the *most* years without an active GOAT candidate (and vice versa for sports with the longest career spans). Nonetheless, when looking at **Figure 2**, one can see that there does not seem to be a relationship between a sport's average career length and the number of years (from 1945 to the present) that a sport does not have an active GOAT candidate. Swimming, which has one of the shortest average career spans (≈ 9.56 years) of the sports that we have included in our analysis, has one of the least years without an active GOAT candidate (8).

Table 1: Number of Years with 0 GOAT Candiates by Sport (since 1945)

Sport	#
Cricket	74
Baseball	38
Golf	21
Hockey	20

Sport	#
Swimming	8
Football	7
Chess	5
Soccer	4
Tennis	4
Basketball	3

While this finding may sound surprising (or at the very least, counterintuitive) on the surface, it can easily be explained. For swimming, we decided that GOAT candidates are those who have held the the world record for the 100-Meter Freestyle race. (While there are many swimming races and events to choose from, this event was chosen because it is often considered to be the most prominent, blue-ribbon event in the sport of swimming.) We also decided to include Michael Phelps due to the fact that he has 23 Olympic gold medals, which is more than twice that of the Olympic swimmer with the second most Olympic gold medals, Mark Spitz (and on top of that, having a list of the best swimmers of all time that did not include Michael Phelps just sounded plain old silly). Using this approach, when diving into the record book, we saw that the world record for said race is broken fairly frequently. Since the world record for the aforementioned race has been broken 34 (**MAKE SURE TO CHANGE THIS TO AN INLINE VALUE)** time since 1945, on average, it has been broken every 2.29 years. But why is this the case? Why swimming world record broken so frequently? Well there are various explanations for these questions as well.

Firstly, we would like to acknowledge that the nature of this GOAT criteria in itself leads to more GOAT candidates. In other sports that we looked at - i.e. baseball, tennis, etc. - to qualify as a GOAT candidate, one would have to exhibit excellence and greatness over a sustained, prolonged period of time. With our criteria for swimming - and running for that matter - this is not the case. One can make it on our list with one good race. In the most extreme case, one can race once, break the world record, and be considered a GOAT candidate.

Swimming is also an interesting case because there have been improvements in equipment that enabled modern swimmers to gain competitive advantages over their predecessors, the most prominent - and extreme - of which were tech suits. Tech suits burst onto the professional swim scene shortly thereafter the 2004 Summer Olympics in Athens when Speedo asked NASA to help it design swim suits will less drag. In the subsequent years, Speedo released the LZR Racer, which furthermore led to the release of various other aerodynamic tech suits - or "susper-suits" - from prominent swimwear companies. During this period - known as the "shiny suit era" - swimming world records were broken at an unprecedented pace. More specifically, *74* world records were broken by swimmers wearing the LZR Racer suit. Consequently, World Aquatics, formerly known as the Federation Internationale de Natation (FINA), controversially banned tech suits in 2009, which has led a noticeable drop-off in the pace at which world records have been broken (McCluskey 2019). Tech suits are a prime example of how advancements in equipment can give athletes advantages over their predecessors. More information about them and their impact(s) on world records can be found here (Filocca 2018).

Contact vs. Non-Contact Sports

One distinction that we chose to investigate is contact sports vs. non-contact sports. An article published by *The American Academy of Pediatrics* titled "Medical Conditions Affecting Sports Participation" divides sports into three main categories: (1) contact, (2) limited-contact, and (3) non contact. Furthermore, contact sports are sub-divided into (1) collision and (2) contact sports. It further states:

"In collision sports (eg, boxing, ice hockey, football, lacrosse, and rodeo), athletes purposely hit or collide with each other or with inanimate objects (including the ground) with great force. In contact sports (eg, basketball and soccer), athletes routinely make contact with each other or with inanimate objects but usually with less force than in collision sports. In limited-contact sports (eg, softball and squash), contact with other athletes or with inanimate objects is infrequent or inadvertent. However, some limited-contact sports (eg, skateboarding) can be as dangerous as collision or contact sports. Even in noncontact sports (eg, power lifting), in which contact is rare and unexpected, serious injuries can occur."

Using the criteria noted above, we decided to classify **basketball, football, hockey, and soccer** as **contact** sports; conversely, **baseball, cricket, golf, and tennis** were deemed **limited contact or non contact**. In doing

so, we were looking to investigate whether the type of sport that a player participates in affects a GOAT candidate's career length.

Finding 3

A Toy Model - Maximal Value for Increasing Distributions

ADD HERE

Filocca, Giulia. 2018. "Can World Records Supersede the Super-Suit Era?" https://www.swimmingworldmagazine.com/news/can-world-records-supersede-the-super-suit-era/; Swimming World.

Gallup. 2017. "Sports." https://news.gallup.com/poll/4735/sports.aspx; Gallup.

McCluskey, Lianne. 2019. "One Decade Later, Do We Miss the Full-Body Tech Suit?" https://www.swimmingworldmagazine.com/news/one-decade-later-do-we-miss-the-full-body-competition-suit/; Swimming World.

Media, News Members, News Freeview. 2017. "The World's Most Watched Sports." https://sportforbusiness.com/the-worlds-most-watched-sports/; Sport for Business.