

Jack Campanella, Connor Paszkiewicz

Sports Analytics Project Report

Professor Barron

11 December 2023

## Age and Performance in the PGA

### **Introduction**

Both being avid golfers, our team decided to make use of a golf dataset on Kaggle which consists of PGA tour data from the last ~10 years. Knowing which sport we were interested in and wanted to perform analysis on, we next considered which specific problem we would like to attempt to answer. Golf is widely known as a sport which can be played well into one's later years of life with little to no issue. This begs the question: what is the true impact of age on performance? Not only does our team want to examine overall how age impacts performance, but we would also like to trace how individual player performance changes over time.

Specifically, we are interested in guiding questions such as:

- When do players typically peak in their golf career?
- When and how severely does a player's golf ability decline with respect to age, if at all?
- What is the overall impact of age vs performance in our golf dataset?

By examining questions such as these, we hope to gain a better understanding of the true impact of age against performance. Is the old saying that "golf knows no age" really true? Through the use of data analysis methods we have learned in sports analytics, we hope to answer this once and for all.

The dataset we have acquired is from Kaggle, although it came originally from <https://www.advancedsportsanalytics.com/pga-raw-data>, which is no longer an active domain. It contains 37 columns including player names, a unique player id, their performance as measured by shots gained metrics for various shot types, and some predictions from Draft Kings (DKP) and Fan Duel (FDP). Each row represents a single players' overall performance at a single tournament. The data goes as far back as 2015 and is as recent as 2022. We have numerical variables tracking players' average shots gained per shot type (off the tee, approach, around the green, putting, overall tee to green, and total), as well as finish position and total strokes to judge a player's performance at any given tournament. In addition to the kaggle dataset, our team used

web scraping techniques to tabulate player age data and season performance metrics from the ESPN website. We then joined our two tables on player name and thus created the completed PGA Tour dataset on which we performed our analyses. Because we are looking at PGA data, it is important to understand that tournaments are broken up into 4 rounds of 18 holes, and players only earn money from a tournament if they make the cut after the second round. Thus, we are interested in analyzing players that make the cut to see what sets them apart from the rest of the field.

### **Related Work**

The first work our team came across which begins to address golfer age and performance is from the Zippia data science team. Zippia has detailed information on golfer demographics, such as age, gender ratio, ethnicity, and salary. Although the work performed by Zippia is rigorous and robust, there is very little analysis on their findings. Rather, Zippia offers up factual statistics with minimal insights. Our work extends beyond simple figures and confronts the question of age versus performance head on. Another study conducted by Richard Schulz and Christine Curnow titled *Peak Performance and Age Among Superathletes: Track and Field, Swimming, Baseball, Tennis, and Golf* hits closer to the mark. In this journal article, Schultz and Curnow find that golfers peak in performance at about 31 years of age with recent data indicating “movement toward younger ages”. Although this study answers the question of the average age at which golfers peak, it was published 25 years ago and fails to address all of our questions. In the journal article *Age and Winning Professional Golf Tournaments*, Gizachew Tiruneh finds that age has a tangible impact on winning professional golf tournaments. That being said, the article was published in 2010 and fails to explore the impact of age on performance beyond mean, median, and mode ages at which professional golfers won various tournaments. We believe that the game of golf, like so many other sports, has changed rapidly over the last few years due to various factors, and thus a more recent analysis is required to identify current trends. Through our team’s analysis, we will utilize the most recent data available to perform a more holistic investigation of age and performance in the PGA as discussed above.

## **Methods**

In terms of our approach to solving this problem, our team first sought to frame the problem at hand. After much discussion, we determined that the most interesting question to answer given our limited time and resources would be the impact of age on performance in the PGA Tour. This question, however, is relatively vague. Accordingly, our team then set out in writing some potential questions we would like to answer which would give us insight into our overarching problem. Apart from the aforementioned questions of interest, such as peak performance age and the extent of the decline in performance as a result of aging, our team also wanted to answer:

- Does a particular group of athletes commonly struggle in a particular area of play?
- Do older players perform better on and around the green than younger players?

Beyond these questions, our team took it upon ourselves to uncover what action an individual player can take to combat any negative impacts of age and performance. For example, if our analysis finds that player performance declines precipitously at 40 years old, what could a 40+ year old player do to avoid this? Should they focus all of their attention on their short game? Should they train more and less frequently? The answers to these kinds of questions are the outputs our team desires from this project. They are outputs that players can feasibly incorporate into their training and play and will provide us the tools required to solve the potential problem of a skill gap, or lack thereof, between age groups of players.

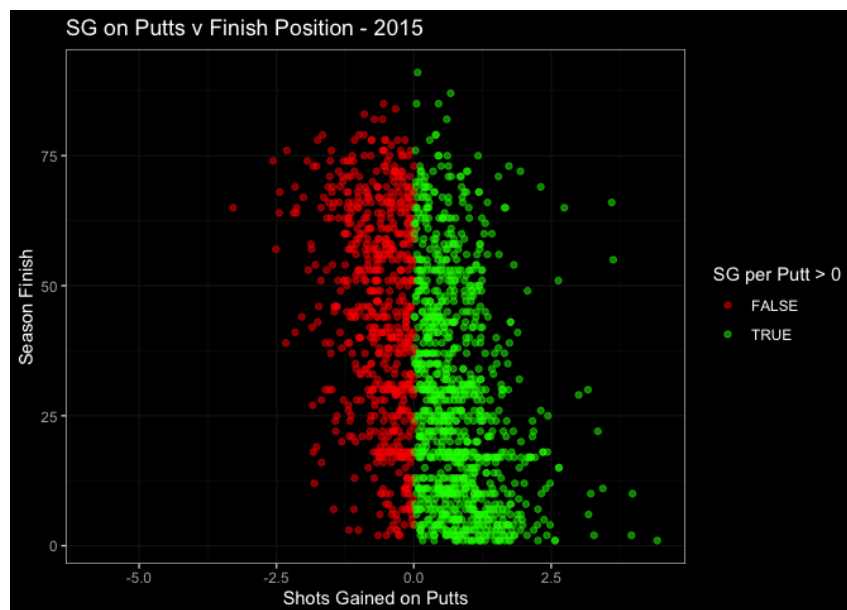
We began our analysis by exploring the tournament-level data from Kaggle. We really had no personal data for the players that we could segment the playing field by, so we looked to ESPN for age data. We had to use web scraping methods to get the list of players and their info for each season, and then combine all of those data points into one table. We then exported that table and loaded it into our analysis code because the web scraping took a few minutes to run so we did not want to run it repeatedly. We did some preliminary data cleaning by changing data types of certain variables and generating a common naming convention that could be used to join our two data tables. We removed any rows that didn't have performance metrics for a player as the focus of our analysis was on using these metrics to compare individuals and groups of players. We found that there were 299 players accounting for 2610 null values, while 354 players accounted for the other 27,272 rows. Thus, we decided to drop these null rows because they

represented a relatively small portion of the data. Once our tables were joined and cleaned, we began our data exploration and analysis.

We analyzed the data by looking at both overall performances from the field of players and from individual players analyses. We began by investigating the importance of putting over the years by visualizing the shots gained from putting metric against overall placement in a tournament. We also analyzed the relationship between age and placement in a tournament through visualizations, and calculated the percentage changes for each of the shots gained metrics from one age group to the next. To see just how strong some of the evidence was, we looked at Bryson DeChambeau, a relatively young player whose play style relies heavily on his shot power rather than technique like putting. Although it must be noted that when we discuss differences between tour pros we are really talking about marginal differences because all aspects of a pro's game must be world-class to be able to compete at such a high level.

## Discussion

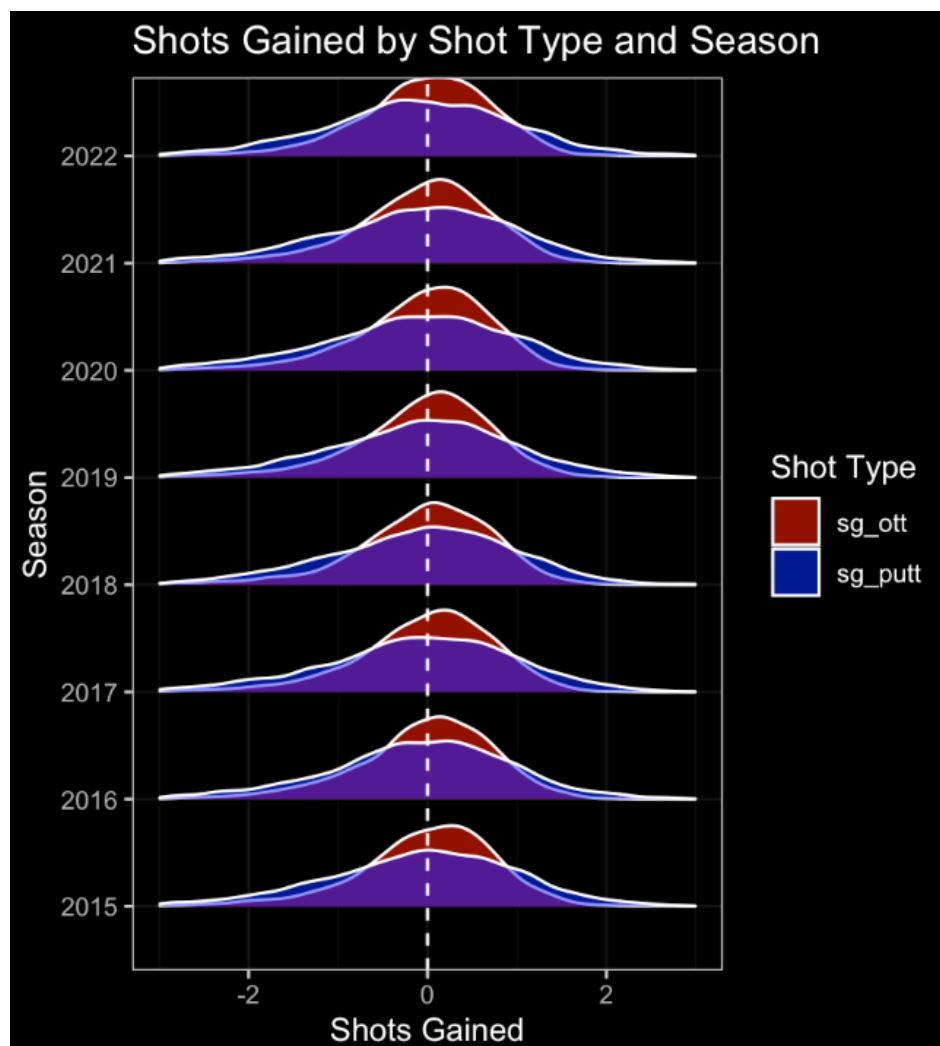
The first analysis we did on our dataset was looking into the importance of putting in the PGA. The PGA has been trying to preserve the skill of reading greens by banning any green-reading instruments. We wanted to see if this rule had any real merit to it or if it was just an attempt to keep the game from becoming too tech-heavy.



From this visualization, we can see that across all years we have data for, the overwhelming majority of players in contention for wins (closer to 1st place at the bottom) are able to putt

better than the rest of the field. Meanwhile players that still make the Friday cut, and thus are better than half of the field already, but still fall well short of first place tend to putt worse than the field. Therefore, we can conclude that putting performance can likely be the difference maker for many tour pros and thus the skill of reading greens is an incredibly valuable asset for players and caddies. We believe this is the case because players can find a wide variety of ways to make it onto the green with different clubs and shot shapes, but once a player is on the green, they face the same challenge as everyone else: the putter.

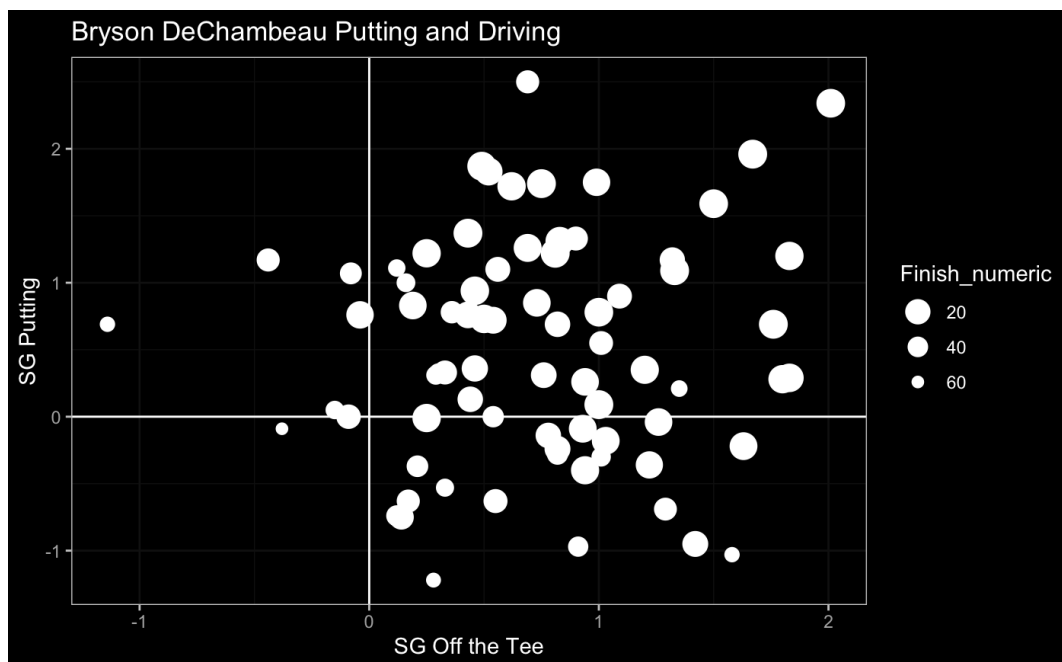
We also wanted to understand where variation in player performance was originating. We did this by analyzing the distributions of strokes gained by different shot types.



We found that shots gained off the tee were concentrated around zero while shots gained in putts had a much wider distribution. Because shots gained off the tee are clustered around zero, we know that players are typically not gaining strokes on the field from the tee box. Instead, players

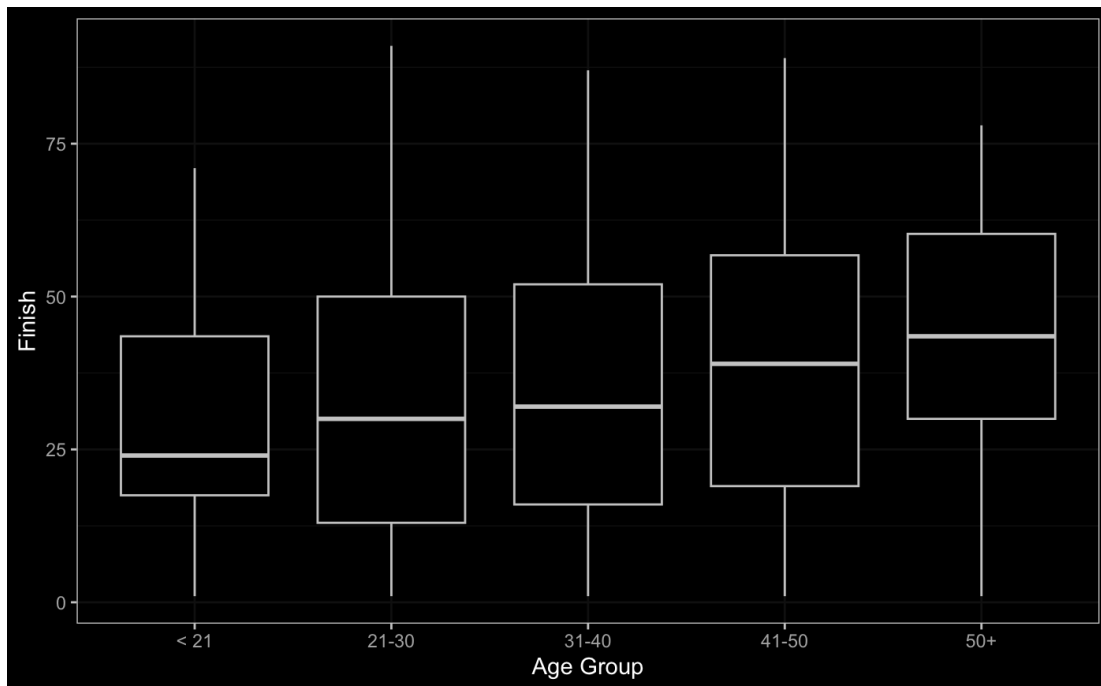
are often outperforming the field, or falling behind, on the greens. This again emphasizes the point that the best players are superior putters and that PGA Tour pros can find numerous ways to make it to the green in the same number of strokes.

We then conducted an individual player analysis on Bryson DeChambeau, a player known for power off the tee rather than his putting. We did this to further explore how significant putting performance is for top player placement on the Tour and to put a potential outlier under the microscope. We compared the impact of putting and driving performance of Bryson on his Tour placements. Our team found that his top placements have highly variable strokes gained from putting, but consistently possess very high strokes gained off the tee. This indicates that Bryson is, in fact, an outlier in the PGA Tour. He is one of the few players who has the luxury of foregoing superior putting performance for extremely strong play off the tee.



Next, we moved into a basic age analysis of finish versus player age group. We split age groups into under 21, 21-30, 31-40, 41-50, and 50+. The data presented a clear distinction between final placements by age group. As the age group increased, the finish position increased each time. At the beginning of this project, we hypothesized that players aged from 31-40 would display the lowest average finish position. We believed this would be the case due to other studies that have been conducted in this area which indicate golfers peak in their performance in their 30s, although an older study we reviewed indicated that there could be a trend towards

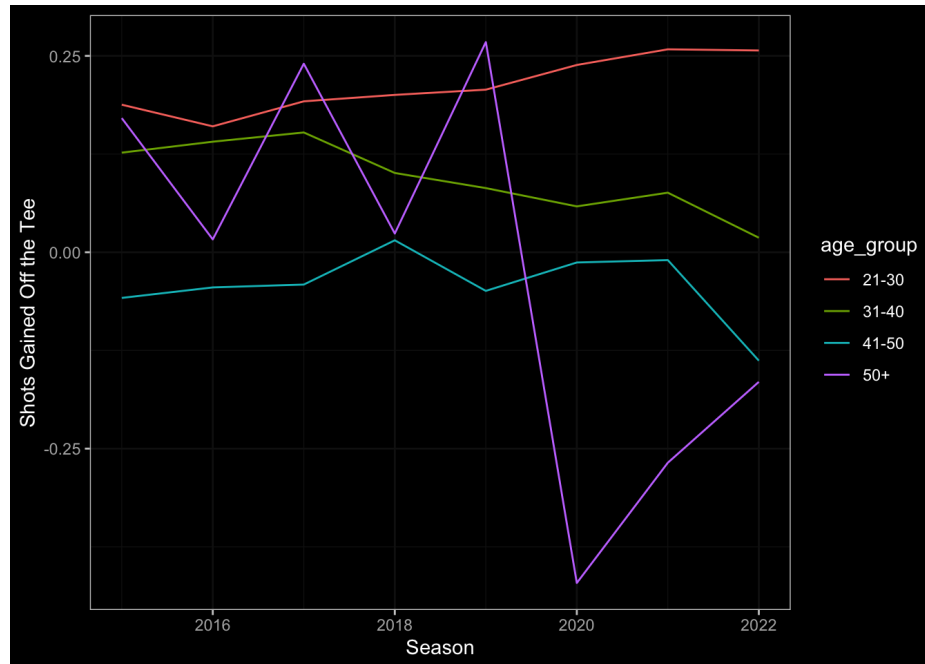
younger players finishing in lower, better positions compared to other age groups. However, due to the less physical nature of golf and the heavier emphasis on mental composure, consistency, and experience, we believed that golf would favor slightly older players. Overall, we were surprised to see how well amateur players were finishing. That being said, we should largely ignore the under 21 group for the purposes of this analysis, as there are only two players who fit this description.



When performing this same age analysis on Bryson DeChambeau, we found that he has been consistently improving as he has aged from ages 23 to 28, with the biggest improvements being from ages 23 to 25. Bryson first played in the PGA Tour at 23 years old. He performed his worst at this age, but has finished at his best at ages 25, 27, and 28. Because Bryson continues to perform at a high level and has not yet passed his peak according to the studies mentioned earlier, we can conclude that Bryson appears to be in line with our holistic age analysis. This is due to the fact that he is performing exceptionally well in that 21-30 group that typically finishes better than all other age groups.

One final question that we wanted to investigate was whether the PGA Tour has actually seen a change in recent years indicating a widening talent gap between younger and older players. We believed this would be most pronounced when looking at shots off the tee because of

the technological improvements in golf balls, swing patterns, and club materials and customizability. Players can tweak an endless number of factors to maximize their ball control and distance which is most exaggerated when hitting with drivers and woods. Therefore, we visualized the average shots gained off the tee for each age group by season, excluding the 2 players under 21 because they did not have enough data points.



This showed us that the 21-30 group has seen steady increases in shots gained from the tee box, while all other age groups have decreased over the same 7-year period. We believe this supports the hypothesis that younger players are indeed getting more of an edge on older players and thus age is playing a bigger role in the PGA now than it did a decade or more ago.

### **Conclusion and Future Work**

As age increases, performance of PGA Tour players decreases. This is an indication that young talent is replacing veterans who have dominated the field for years, and is likely due to the fact that golfers are being introduced to the sport at increasingly younger ages with increasingly more resources to improve their game. Training methods have taken significant strides in recent decades as golfers and coaches alike utilize data analytics to optimize their training. We also know that, overall, putting skill is crucial and very few players win while putting worse than the field. That being said, there is no one way to win. Some players, such as Bryson DeChambeau, can specialize in certain shot types to gain a unique advantage on the field. The next step for this



project, given more time and resources, would be to evaluate the impact of training on performance. If we were able to collect more granular data on which players are using highly-optimized training regimen and which are using more traditional methods, then we could compare the performance of each group to uncover if there is a significant impact on performance for different training techniques, and perhaps develop combinations of training methods that could maximize all aspects of a player's game. We could also discover if differences in training is what truly separates younger and older players, or if there is some other factor like rising competition levels in youth sports.

### **Contributions**

Both team members contributed equally to all parts of the project final report, presentation, data gathering and analysis.

## **Bibliography**

1. “2022-23 PGA TOUR Stats.” ESPN, [www.espn.com/golf/stats/player](http://www.espn.com/golf/stats/player).
2. Fried, Harold O., and Loren W. Tauer. “The Impact of Age on the Ability to Perform under Pressure: Golfers on the PGA Tour.” *Journal of Productivity Analysis*, vol. 35, no. 1, 16 Sept. 2009, pp. 75–84, <https://doi.org/10.1007/s11123-009-0151-9>. Accessed 11 Nov. 2019.
3. “Golf Professional Demographics and Statistics [2021]: Number of Golf Professionals in the US.” [Www.zippia.com](http://www.zippia.com), 29 Jan. 2021, [www.zippia.com/golf-professional-jobs/demographics/](http://www.zippia.com/golf-professional-jobs/demographics/).
4. Hayslip, Bert, and Trent A. Petrie. “Age, Psychological Skills, and Golf Performance: A Prospective Investigation.” *The Journals of Gerontology: Series B*, vol. 69, no. 2, 1 Mar. 2014, pp. 245–249, [academic.oup.com/psychsocgerontology/article/69/2/245/545993](http://academic.oup.com/psychsocgerontology/article/69/2/245/545993), <https://doi.org/10.1093/geronb/gbt010>. Accessed 25 Sept. 2020.
5. “PGA Tour Golf Data - (2015-2022).” [Www.kaggle.com](http://www.kaggle.com), [www.kaggle.com/datasets/robikscube/pga-tour-golf-data-20152022/data](http://www.kaggle.com/datasets/robikscube/pga-tour-golf-data-20152022/data). Accessed 11 Dec. 2023.
6. Schulz, R., and C. Curnow. “Peak Performance and Age among Superathletes: Track and Field, Swimming, Baseball, Tennis, and Golf.” *Journal of Gerontology*, vol. 43, no. 5, 1 Sept. 1988, pp. P113–P120, <https://doi.org/10.1093/geronj/43.5.p113>.
7. Tiruneh, Gizachew. “Age and Winning Professional Golf Tournaments.” *Journal of Quantitative Analysis in Sports*, vol. 6, no. 1, 28 Jan. 2010, <https://doi.org/10.2202/1559-0410.1209>.