MONTANA STATE UNIVERSITY

STATISTICS BEHIND THE SKILL: CLUSTER ANALYSIS AND DATA VISUALIZATION ON DISC GOLF DATA

July 13, 2018

Elijah S. Meyer

Department of Mathematical Sciences

A writing project submitted in partial fulfillment of the requirements for the degree

Master of Science in Statistics

APPROVAL

of a writing project submitted by
Elijah S. Meyer

This writing project has been read by the writing project advisor and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Statistics Faculty.

Approved: ______________________________

Dr. Jennifer L. Green

Writing Project Advisor

Date ______________________

Approved: ______________________________

Dr. Mark C. Greenwood

Writing Projects Coordinator

Date ______________________

Contents

Abstract

Disc golf, a game of strength, strategy and control, is a competitive sport with the goal of throwing a plastic disc into a metal cage. The objective of this research was to develop data-informed preparation strategies for this sport using performance indicators from elite disc golf players. This was explored through two different studies. For the first study, 3D visualizations of disc golf holes using latitude, longitude and estimated elevations from Google Earth were created to emphasize characteristics that still pictures fail to capture. For each hole, the 3D visualizations and still pictures were both presented to disc golf players. Interviews conducted on characteristics noticed and players' preparation techniques when observing a still image vs. a 3D visualization revealed that players assess the shape and elevation change of the land more effectively with the 3D visualization than with a still picture.

In the second study, hierarchical clustering methods were performed on a suite of covariates measured over the 2016 and 2017 disc golf seasons to identify important factors associated with player performance. These results suggest that, although putting is an important part of the game, it does not describe differences in professional players' performance. Separation of placements between professional players may be due to performance off the tee.

Together, these studies help to advance overall preparation strategies for players in this growing sport. The information provided from both studies can help players understand the factors that separate professional level players' performance, providing insight into what facets of the game an individual may want to prioritize. Furthermore, with the additional information 3D visualizations provide, players can better understand the task at hand in order to optimize their performance and keep their scores low.

1. INTRODUCTION

1.1 Background

Disc golf, originating in 1976 (Disc Golf Association, 2017), is a sport yielding a common goal across players to throw a variety of plastic discs into a metal cage in as few throws as possible. With over 28,000 members of the Professional Disc Golf Association in the United States alone (Professional Disc Golf Association, 2018), disc golf has a growing community of players who compete across over 2,000 worldwide events a year.

The professionals of any sport can act as a staple to model your game after. Players of all skill levels can observe professionals in an attempt to improve one's own game. This can be done by attending a live event or by watching previously recorded footage on the internet of a disc golf tournament. In addition, players can obtain the score card for a tournament, but this only gives the total score of each competitor. For example, Ricky Wysocki won the 2017 Aussie Open with a combined score of 47 under par (-47). Anyone can identify that this is a fantastic score, but it is hard to quantify what factors led to this score. How well did he putt? How well did he drive? What did he do better than the others to win?

In disc golf, knowing what factors contribute to a professional's winning score can be useful information to improve one's own game. It's also advantageous to know information about the courses on which you will be competing. Often, individuals use pictures of unfamiliar holes, or holes to be played in competition, to help prepare and familiarize themselves for the upcoming play. This type of preparation is a useful tool for disc golf players. However, still 2D pictures vary in quality and provide limited information about a given hole. In general, there is a need for better visualizations, and a data-informed understanding of what indicators impact performance in order to help players enhance their preparation strategies.

1.2 Motivation

Often monetary gain can motivate competitive disc golf players to practice and refine their game in order to place higher in the standings. Figure 1 shows the difference in earnings among professional players, along with their rank at the end of the 2017 season (Professional Disc Golf Association, 2018).

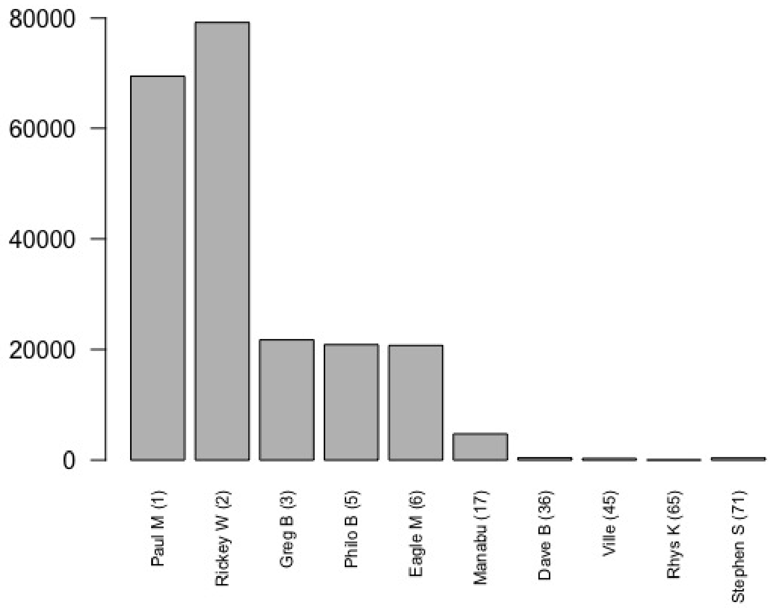


Figure 1: Players' earnings and ranks in the 2017 disc golf season

Although these numbers are not as high as other sports such as the National Football League, there is a clear difference in earnings between higher and lower-ranked professionals. In addition, monetary gain isn't for professionals only. There are many tournaments worldwide that offer cash prizes to those who throw the lowest total score, emphasizing that monetary incentive spans across a lot of players in a variety of locations.

Money can fuel the fire of competition for individuals. However, the desire to improve is not driven solely by monetary gain. It may also be fueled by a passion for competition and/or personal goals. Regardless of the reason to get better, it is a common theme that the majority of players strive to improve every time they step onto the course or practice field. This further expands the player pool that could benefit from novel visualizations and data–based information about the sport that are currently lacking.

1.3 Research Questions

This research encompasses two different studies. Study 1 provides a novel way to prepare for and analyze different disc golf holes prior to playing them using 3D visualizations that are currently not being used in today's sport. We describe the generation of 3D visualizations that allow players to rotate the landscape and assess different angles/lines that still pictures fail to provide. We then compare the characteristics disc golf players emphasize across these different visualizations in order to address the following research question: What types of information do different visualizations emphasize and how does that reflect in players' strategies?

Study 2 focuses on different skillsets measured during disc golf world tour events and what skills separate groups of professional players (elite tier, mid tier, low tier) from one another. This study addresses the following research question: Which skillsets at the professional level differentiate players' performance?

In the following sections, we describe each of the studies and discuss how the results are applicable to someone interested in competing in the sport of disc golf. Together, these studies offer useful information for preparing strategies to implement and for enhancing one's overall game.

2. Study 1: Picture Perfect Preparation

The objective of all disc golf players is to complete a course in as few throws as possible, so players must perform to the best of their abilities on each hole. With only a few strokes separating players' scores, each hole matters; a hole that was not performed well on could be the difference between placing high, or not at all.

Prior preparation plays a key role in performance. In any context, the more you prepare for the current task at hand, the more likely you are to perform better at that task. The same rules apply for disc golf. Currently, players use a variety of preparation techniques to help improve their overall game and lower their scores. If a player has previous experience playing a course, or has time to walk the course prior to playing, the information they recall and collect can help provide information on a variety of aspects and help generate ideas for potential strategies to be implemented. However, with tournament directors potentially changing the tee pad and basket locations between tournament visits and/or between rounds played, previous recollection may not provide the information necessary to perform well. Furthermore, it may not be feasible to incorporate the amount of time needed to walk an entire course prior to playing, and even if so, the mental snapshots may not provide the full story of each hole across the course.

A picture is worth a thousand words. Is it worth using for preparation? Still pictures provide information about a hole that an individual can study remotely before a competition. These pictures can be found on a variety of websites, along with the length and par of a hole. However, depending on a picture's quality and the features it captures, studying still pictures may be a frustrating and misleading experience come time to play. This highlights a need to create additional tools that help emphasize different characteristics of a hole and provide players the opportunity to develop strategies to optimize their performance.

In this section we present a novel tool we created for disc golf players to better visualize holes. We first describe the methodology used to create the 3D visualizations and then share our research study to explore their use in players' preparation. We conclude with a summary of our research findings.

2.1 Methodology

Three disc golf holes from courses located in Bozeman, Great Falls and Helena, Montana were selected for this study. These holes were strategically selected from courses that we had experience playing so that we could confirm when the 3D visualizations were accurate representations of each hole. Still pictures associated with each of the three holes were obtained from dgreview.com (Becker, 2014; Danger, 2010), a popular website for players to access a variety of information on courses throughout the world. Players often utilize the pictures and information presented on this website before playing, so these pictures represent standard resources players access prior to competition.

2.1.1 Development of 3D Visualizations

The process of forming the 3D visualizations started with the use of Google Earth. Google Earth, a computer program that renders a 3D representation of Earth based on satellite imagery (Google Earth Pro, 2018a; 2018b; 2018c), was used to collect latitude and longitude coordinates corresponding to the landscape of a disc golf hole by placing pins along the imagery. Elevations were estimated from these coordinates using the "googleway" package in R (Cooley, 2018). The landscape chosen to be modeled was based off importance, defined as where the average disc golf player would have their disc make contact with the ground off the drive. Landscape containing obstacles for the player to consider, and general changes in landscape that may be useful for the player to know about during the preparation phase were also captured by pins. The pins were not systematically placed. The landscape was exhausted, meaning that the landscaped deemed important was covered by many different pins. However, modeling each and every point across the landscape wasn't feasible using this point and click technique. To capture the missing elevation points needed to construct the image, universal kriging was performed.

Universal kriging is a spatial interpolation method based on statistical models that account for the autocorrelation between each of the estimated elevation points and their location (Ersi, 2017). Universal kriging is used to interpolate geostatistical data across a spatial domain using a specified parametric spatial correlation structure; this covariance function, typically specified in terms of the semivariance between observations, is characterized by three key parameters: (1) the nugget effect, (2) the range, and (3) the (partial) sill. The nugget effect describes the semivariance between two observations arbitrarily close together. The range parameter describes the distance at which the spatial correlation has little to no effect on the response; practically, it is often thought of as the point at which the spatial autocorrelation dies off. And finally, the sill parameter describes the semivariance of the process once the range has been attained; it can be thought of as a scaled form of the natural variability in the process when observations are sufficiently far apart.

The exact relationship between these three parameters and the response depends upon the family from which the semivariogram is defined. In the context of predicting elevation, an exponential covariance function is a reasonable choice, as we expect the correlation between elevations to die off as the distance between two points increases. Other covariance functions were considered and may also be reasonable for this type of process. However, with the initial justification of an exponential covariance function coupled with the end goal of building the picture, not inference, other functions were disregarded.

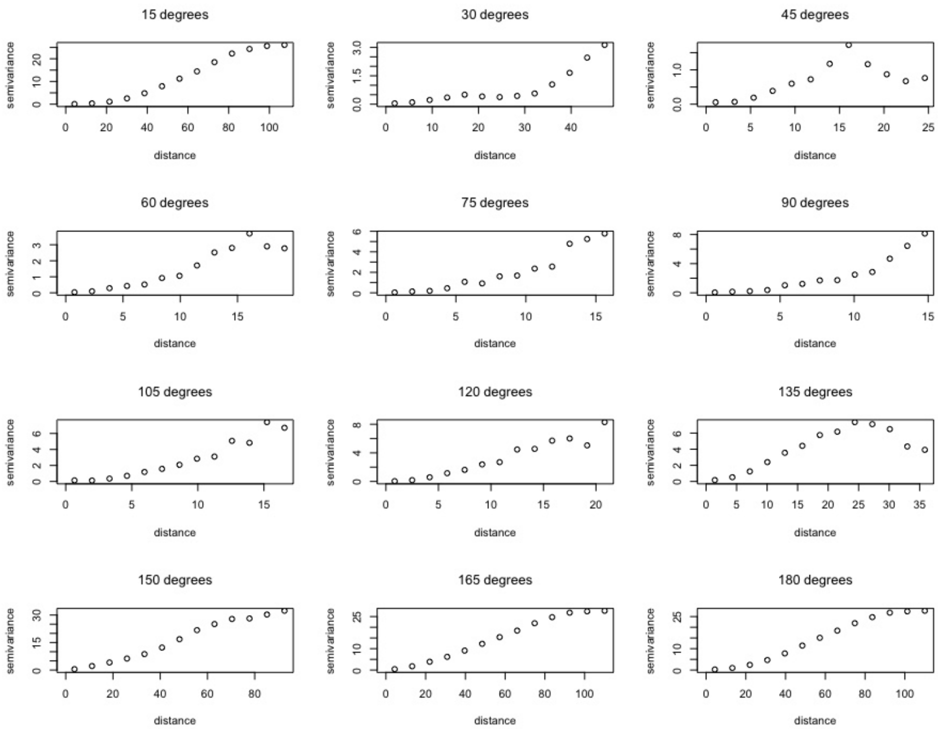


Figure 2: Semivariograms based on different degrees for the Great Falls hole

If the landscape being modeled resembled a symmetric hill on all sides that peaked in the middle, then the relationship between the top of the hill and the bottom of the hill going East would be the same as the relationship between the top of the hill and the bottom going West. This would imply that there was no directional dependency between elevation points. However, upon analysis of the semivariogram, the range was never attained under this assumption, providing evidence that anisotropy may exist. There was evidence that the correlation between points may depend on direction. To account for this, the direction of minimum variation was estimated by using different degrees. The degree that showed the most well-behaved semivariogram (i.e., the semivariogram that showed evidence of attaining the range) was used. For example, as shown in Figure 2, twelve different semivariograms at varying degrees were explored. Semivariograms, such as the one at 30 degrees and the one at 75 degrees, did not show evidence of attaining the range, and were not considered of use.

The semivariogram at 60 degrees flattened out after a distance of 15, providing evidence that, at this degree, the range was obtained. Therefore, the model used to estimate elevations around our collected data was fit with a degree of 60, a range of 15, and a sill of 1.5. Figure 3 displays the chosen empirical semivariogram for each hole's model with the corresponding degree, range and sill value.

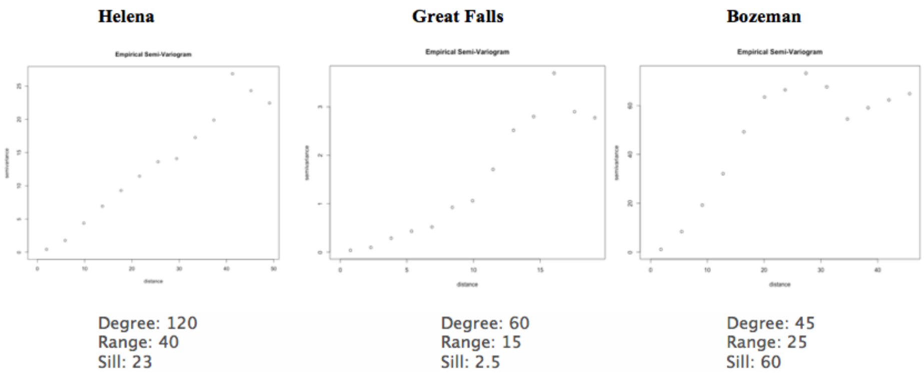


Figure 3: Parameter values for each of the three models
(a) Helena hole (Left) | **(b)** Great Falls hole (Center) | **(c)** Bozeman hole (Right)

A grid of observed latitudes (x), longitudes (y), and elevations (z), as well as estimated values created using universal kriging, were inputted into functions contained in the plot3D package in R (Soetaert, 2017). These values were used to create 3D visualizations of the three disc golf holes chosen for this research.

2.1.2 Survey Development

Survey questions for the imagery were developed to have players articulate what characteristics they were observing and how they planned to play the hole they were assigned. Players' responses were based off information they observed through either the still imagery or the 3D visualizations. The questions were constructed to elicit rich information about players' observations and strategies, and to discourage "generic" responses such as, "I want the disc to be in the basket," or "I always want my disc to land 15 feet from the basket." The questions prompted players to think critically about the information they could observe, carefully instructing them to only use the information they could see from a given image (still or 3D). The same questions were used for both sets of imagery. These survey questions can be found in Appendix A.

2.1.3 Participant Selection

Participants for this study were recruited from a variety of disc golf communities including Great Falls Disc Golf Club, Headwaters Disc Golf Club, etc. Players associated with these established communities have experience playing disc golf and are knowledgeable about playing tournaments. This is crucial as the survey questions were tailored to these type of competitors.

Players from these communities were initially contacted via a Facebook group message. Eleven of the players agreed to participate, and each was assigned a hole based on prior experience. Players that have played a hole multiple times may tend to disregard information from both sets of imagery, and may instead rely on prior experience to answer questions about that hole. This form of preparation was not the focus of this research, so each player was assigned a hole with which they were unfamiliar. Players who were unfamiliar with more than one hole in the study were offered to answer questions about a second hole. As a result, five players were assigned to the Great Falls hole, four players were assigned to the Bozeman hole, and four players were assigned to the Helena hole.

2.1.4 Data Collection

Players who agreed to participate in this research were sent the still pictures, survey questions, and detailed instructions for their assigned hole(s) via email. This email also contained a video link providing access to the 3D visualization. Players were instructed to only use the information from the still pictures when filling out the first survey. Players had the opportunity to observe the still imagery for as long as they wanted and could reference the pictures while answering the survey questions. The still pictures and survey quesions can be found in Appendix A. Hole distance and par was also provided in the top left corner of the survey.

After answering the questions with the still pictures, participants then answered the same set of questions using only the information provided by the 3D visualization. Players participating in the survey were shown the 3D rendition of the hole via a video link (see Appendix B). Each video simulated a player moving the rendition of the hole in order to observe different angles and obtain information from all sides of the visualization. Players were allowed to watch the video multiple times.

There was no time constraint on the completion of the questions; at their earliest convenience, players returned an electronic file with their completed responses via email. Throughout the process, players were allowed to ask clarification questions, but no information regarding the hole that would alter characteristics noticed or influence strategies developed was given during the study.

2.2 Analysis

To identify observable characteristics based on still imagery, survey responses for the first question were organized by hole. For each hole, all of the responses were read and codes were constructed to describe the characteristics noticed in the still image(s). Codes were constructed separately for each of the three holes. Once these codes were identified for each hole, the codes across players' responses for a given hole were analyzed to identify repeated use or mention of distinguishable patterns or phrases. For example, for the Helena hole, players identified "trees" and "bushes" multiple times; these codes may represent objects to avoid, potential landmarks to throw next to, etc. Next, these codes and patterns were analyzed across the three holes to compare and identify common themes of characteristics players noticed. Codes that shared similar information supported an overall theme. Codes that were similar to one another, but dissimilar to others, provided evidence of an additional theme.

A similar process was carried out for responses based on the 3D visualizations. Like before, players' responses to the first question were analyzed to assess noticeable characteristics. The themes that emerged from the codes identified from the responses associated with both the still and 3D image were validated by cross checking them with the original codes and survey responses. During this process, we specifically looked for counterexamples that led to the identification of other noted characteristics that contradicted or were not captured by the current themes. None were observed for these data.

Next, we identified the strategies players developed by analyzing their responses to the second through fourth survey questions. These three questions targeted the identification of play style brought about by different imagery. Play style, here, is defined as how a player logically thinks through the different aspects of playing a hole. This includes where shots will land, what obstacles should be avoided, etc.

For each type of imagery (still, 3D), play styles were constructed based on the responses players provided. Codes were assigned to each player's response for each of the three questions for both sets of imagery. Once these codes were identified, they were analyzed across players and across questions for a given set of images (still or 3D). For example, a player identifying changing landscape that they would like to avoid for one question, and another player identifying a distinct place they would like their disc to land because of the changing landscape for another question were both assigned codes such as "landscape" and "elevation." These codes were then compared across players' responses to different questions for a given image. Codes that shared similar characteristics provided evidence of emerging themes describing play style. These themes were used to characterize and compare the different types of play styles across the two sets of imagery, and they were summarized for each of the three holes in the study.

Finally, we investigated the relationship between the characteristics emphasized by a given set of images (still or 3D) and the resulting play styles. For each type of image, this relationship was explored by identifying the occurrence of codes and themes associated with the first survey question within the play style themes constructed from players' responses to Questions 2, 3 and 4. These relationships were then compared across the two types of images. Together, these analyses provided information about the differences in characteristics noticed, play style, and their relationships across the different images.

2.3 Results

In the following sections, we describe the characteristics observed for each type of imagery as well as how these characteristics appeared in the strategies developed. We also compare players' strategies across the different types of imagery for each hole.

2.3.1 Characteristics Observed

Characteristics observed and documented from participants, regardless of hole, were noticeably different between the still pictures and the 3D visualizations. Objects, such as trees and bushes, were dominant themes when players were observing still pictures. When asked to identify characteristics they noticed, participants responded by stating they noticed "no physical objects such as trees" for the Bozeman hole, and "a lot of trees surrounding the area I would like to land in" for the Great Falls hole. Across the holes, participants consistently noted these objects (or the lack of objects) when looking at the still pictures. With the objective being to make your disc into the basket in as few throws as possible, it makes sense to take note of potential obstructions. Furthermore, the still pictures captured the sizes and circumferences of trees, providing players useful information about these objects when playing.

When looking at still pictures, players were also more apt to identify textures associated with the land their disc may skip on. For example, one participant wrote, "Grass seems to be thick enough," indicating that a disc may not have much reaction when hitting the ground. Pictures, with the ability to show detailed features, emphasize the "feel" of the landscape, which was picked up on and identified by players in this study.

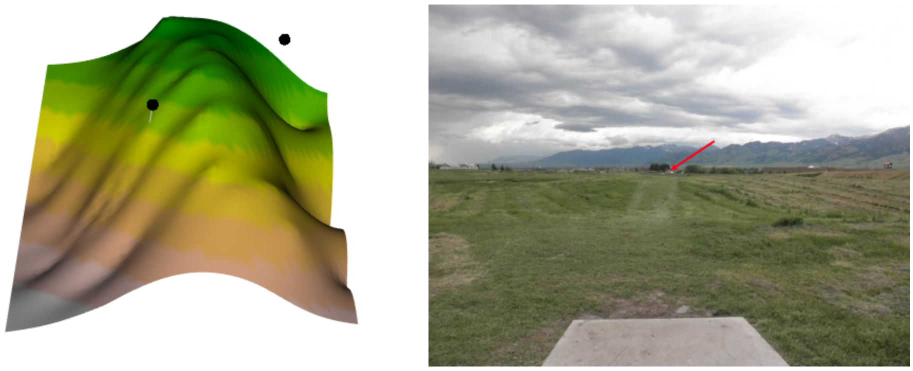
Participants noticed different characteristics when they viewed the 3D visualizations. The 3D visualizations do not have objects to scale and do not accurately capture the texture of the land. However, "change in elevation is very apparent." With the ability to rotate the constructed image, players were able to get a better sense of the entirety of a hole and identify the shape and elevation change around a basket. This was a common theme across participants and holes. Noticing "concave areas" and "fairly steep slopes" were riddled among responses, with little to no mention of the objects identified with the still pictures. This isn't to say objects were not mentioned at all, but they were observed in a very different manner. The rotational ability of the 3D image allowed players a top down view of the hole. No longer were they assessing the objects they were trying to avoid. Instead they were characterizing potential paths based on density of the object because the "trees are now countable."

Still Images	3D Visualizations
Objects	Elevation Change
Texture	Top Down View

Table 1: Themes of Characteristics Noticed

2.3.2 Players' Strategies

The different sets of imagery honed players in on different characteristics, which appeared throughout their strategies. When identifying a strategy for a given hole, players used the characteristics emphasized by the type of image (still vs 3D) they were observing. The Bozeman hole (Figure 4) was categorized as a hole that may not be much of a challenge. Given the information perceived from the still picture, with "no objects" identifiable, players were trying to throw as close to the hole as possible without regard for any potential threats. "The hole seem[ing] ... straight and flat...with lack of trees" gave players confidence that any shot was appropriate and any landing spot would suffice. Landing "straight ahead" from the basket appeared to be a non-issue: "From what I can see that would be just as good a place from 30 feet as any." Players looked for objects (a theme identified above) and discarded reservation about these objects when none were observed.

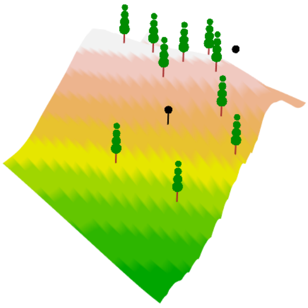


(a) 3D visualization of the Bozeman hole **(b)** Still image of the Bozeman hole

Figure 4: Images observed by players for the Bozeman hole

When observing the 3D visualization of the Bozeman hole, players changed their strategies. The players prioritized where their discs landed and laid to rest instead of writing off the hole as easy. For example, one player noted that he wanted his disc to land "below the basket to give...[himself] an uphill putt." This was a stark contrast to what was strategized when using only the still image. With the change of elevation around the basket and the steep nature of the surrounding landscape, players used this new set of information to come up with a different strategy to optimize performance.

When looking at the still imagery for the Great Falls hole (Figure 5), every participant based his strategy on the trees' locations. Whether players wanted to "avoid having ...[their] disc hit an early tree" or they wanted to throw "towards the third pine tree," the bulk of each player's strategy revolved around object location, a theme emphasized across each hole. However, this quickly shifted when given only the 3D visualization. Rotation of the hole allowed players to view the steep hillside, with a player noting that his "disc may end up at the bottom of the hill," something he was trying to avoid. The change in landscape that the basket sits on was now the focus of concern, with strategies detailing how each individual would prepare for the elevation change. Objects were also used in a different light. With the ability to "count trees with confidence and plan more exact," players had a new way to identify which line to throw off the tee. The focus was shifted from avoiding the trees to counting them and creating a path, a strategy arising from the top down approach unique to the 3D visualizations.



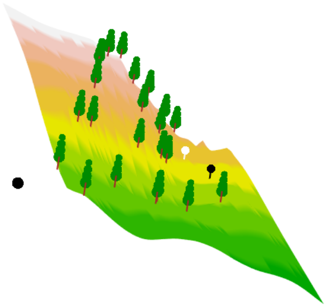
(a) 3D visualization of the Great Falls hole

(b) Still image of the Great Falls hole

Figure 5: Images observed by players for the Great Falls hole

The Helena hole (Figure 6) had similar results as the Great Falls hole. Characteristics players noticed in each of the images informed their strategies, providing obvious differences in play style. When studying the still image, one player indicated that he would "hyzer around the left side of the left tree. Trying to avoid all of the trees." This quote clearly indicates the fixation on trees. Another player noticed a similar line wanting "to avoid the tree to the right of [him]" hoping to "land in the shrubby bushes" so his disc ended up close to the basket. In general, trees and bushes (objects that are emphasized in these pictures) were the main focus of the strategies players developed when only using still imagery. However, players were left wanting more information. One player was left wondering "just how many trees are in my way." Another inquired about "trees further up the fairway on the left that ...[his] preferred discs flight might come in contact with."

With the 3D visualization having the capability of being rotated, the top down view prescribed unique information, clearing up the questions associated with field of vision. Based on this information, players provided more detailed strategies; a player now wanted the disc to land "at the bottom of the concave area ... or just slightly up the slope on the left" instead of identifying objects to land by. Much like in previous holes, players based their strategies on the information emphasized in the 3D visualizations. With this information, players noticed "the fairly steep slope left to right" and used this to identify that they wanted their disc to land "on the side of the hill and possibly not roll down." Overall, disc golf players' strategies changed when given a set of different imagery.



(a) 3D visualization of the Helena hole

(b) Still image of the Helena hole

Figure 6: Images observed by players for the Helena hole

2.4 Summary

Based on the stark contrast in characteristics noticed across the different images by this sample of disc golf players, the 3D visualizations appear to allow players to identify and study different aspects of a disc golf hole. These differing characteristics were reflected in players' strategies. The still pictures emphasized objects and textures, leading players to use information about trees, bushes and thickness of grass to develop a strategy. Players using the 3D visualizations took advantage of the new information about slope and elevation change, developing strategies to handle these challenges.

This study offered information about how players can use different visualizations to develop strategies for playing a hole. In the next study, we explore what we can learn from professional players, and how this can help one's own game.

3. Study 2: Professional Help

Players have the ability to use a wide variety of tools to help refine their game. One common technique, used across a wide variety of sports, is to look up to who is competing at the highest level. Professionals can often be a model for amateur players to mimic or learn from. In disc golf, this is often done by analyzing score cards, admiring the low scores professionals shoot. However, score totals provide little information about how that score was obtained. One could watch film of a given tournament, but it's difficult to identify what aspects of a player's performance are most important. It may be obvious to some who the "good" players in the sport are, but why are they considered good?

There is a natural hierarchy across the population of professional players. Players that show off a superior skill set and are consistently performing at a high level are considered to be the players at the top of their game; these players are elite. Players that show inconsistencies and poor rankings are considered not as good (low tier). The rest of the players encompass the middle ground (mid tier). However, inclusion of players in these tiers is not well defined. Without an easy way to assess what skills are on display with professionals, this study seeks to identify what skillsets differentiate tiers of professional players.

3.1 Methodology

In this section, we will discuss how professional disc golf players were selected and the variables measured on each player during tournament play. We will also explore the methodology used to cluster professional disc golf players based off metrics recorded during play across two different seasons with a focus on the 2017 season.

3.1.1 Data

Performance data from each round of the Disc Golf World Tour Series in 2016 and 2017 was scraped from their official website (Spin18, 2016, 2017a). For a given season, players who had played at least four rounds were included in the data set. For each player, we collected round information on the following variables: Green Hits Percentage (GRH.), Inside Circle Putting Percentage (ICP.), Bullseye Hits Percentage (BUE.), Outside Circle Putts Average Per Round (OCPavg) and Penalties per Round (PenaltiesPerR). Descriptions of these variables are provided in Table 2. Penalties per round, a variable not readily available on the website, was created by taking the number of penalties each player threw divided by the number of rounds the player had played throughout the season.

Each of these variables was considered as each provides unique information that is recognizable and easily relatable to the majority of disc golf players. The variables span the majority of skillsets on display during competition, providing insight on a variety of factors that make up a player's game from how accurate they putt (ICP., OCPavg) to how consistent they are off the tee (GRH., BUE.).

Variable	Description
Green Hits Percentage (GRH.)	The percent of times a player's disc lands within a 10 meter circle around the basket across rounds.
Inside Circle Putting Percentage (ICP.)	When a disc lands in the basket, it is considered a success. ICP. looks at a player's percentage of successful amount of putts made within a 10 meter circle.
Bullseye Hits Percentage (BUE.)	The percent of times a player's disc lands within a 3 meter circle around the basket across rounds.
Outside Circle Putts Average Per Round (OCPavg)	The average number of times a player makes a shot outside the 10 meter circle (green) per round.
Penalties per Round (PenaltiesPerR)	The average number of times a player throws their disc in an illegal area, causing them to take an extra stroke for that hole.

Table 2: Variable Descriptions for Study 2

3.1.2 Analysis

A cluster analysis was implemented to find distinct groups of players that show similar characteristics in their performance. In order to carry out the cluster analysis, we first identified a distance metric. The further players are from each other based on the measured variables, the less likely they will be grouped together in the same tier. Euclidean distance, a common distance metric that is the default in the dist() function in R was used in this research. This metric is the straight line distance between two points in Euclidean space. In general, this distance between two points is calculated by

$$d(\mathbf{p},\mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + ... + (p_n - q_n)^2} \tag{1}$$

where **p** and **q** are vectors of information on measured variables between two players.

Once a distance metric was decided upon, we chose a method to form the clusters. In this study, we used Ward's method, which minimizes the error sums of squares (ESS) within K clusters.

The ESS calculation for a single group is formed by

$$\sum_{i=1}^{n} \mathbf{X_i}^2 - \frac{1}{n}\sum_{i=1}^{n} \mathbf{X_i}^2 \tag{2}$$

where $\mathbf{X_i}$ is a vector of measurements for the ith player and n is the total number of players in the study. However, as more clusters are considered, the total ESS gets partitioned into group ESS.

$$ESS = ESS_{group1} + ESS_{group2} + ... + ESS_{groupK} \tag{3}$$

At each step, the union of every possible pair of clusters is considered, and the merger that results in the smallest increase in ESS for group K is used. This means that players who have similar statistics across the set of variables, or minimal distance away from each other, will be more likely to be in the same cluster.

Dendrograms, common visual tools for cluster analysis, are tree-based displays of hierarchical clustering results that show how each observation or cluster of observations is merged together, and how the final clusters are formed. The merging of these clusters can be expressed visually by the lines in a dendrogram that connect multiple players together as you move up the tree. A line is then drawn horizontally across the dendrogram that cuts the sub trees into groups. These groups represent the final clusters. The numbers within each cluster represent players for that season. A cut too low will result in many and potentially uninformative clusters of players. A cut too high might not tell the full story. In this study, we considered the sizes of the changes in cut heights; often, a "large" change indicates the appropriate number of clusters to interpret (Everitt & Hothorn, 2011). We also used intuition about the general hierarchy of tiers assumed among the population of disc golf players to help justify the number of clusters chosen.

3.2 Results

When examining the dendrogram constructed using the 2017 season data (Figure 7), we observed a relatively small change in height between three to four clusters, and a large change between four to five clusters. This suggests four clusters would be appropriate to interpret. However, with previous justification for three tiers of players as well as the small difference in cut height between three and four clusters, we proceeded using three clusters. The 2016 dendrogram (see Appendix C) yielded similar results as above.

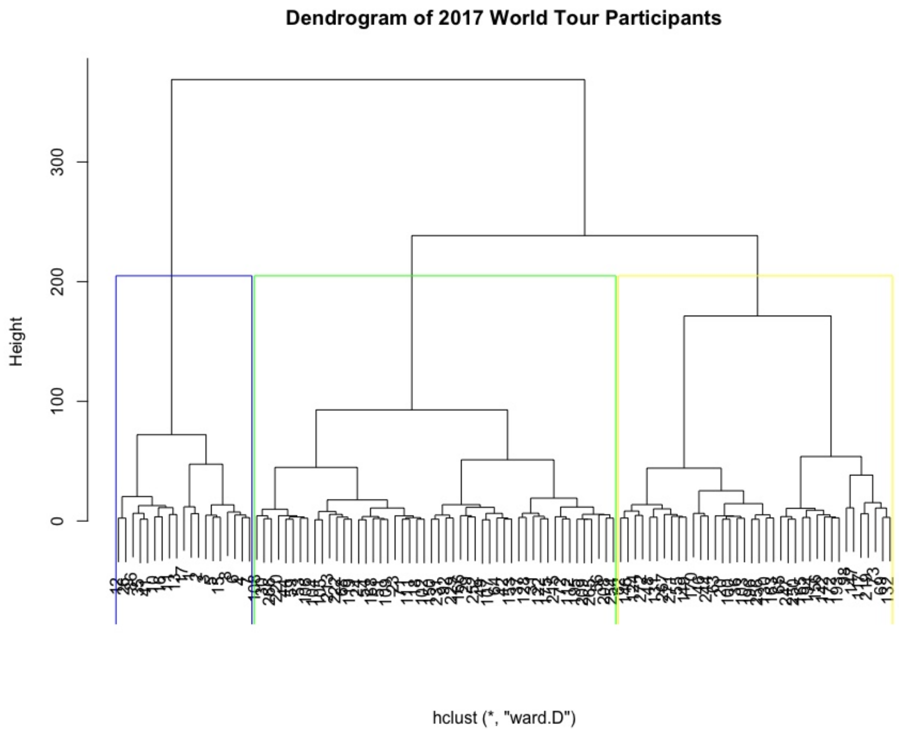


Figure 7: 2017 Disc Golf World Tour Series dendrogram with three clusters

Based on the reputation, skillset and credentials of the players within Cluster 1, there is strong evidence to conclude that Cluster 1 is composed of elite tier players, a tier we suspected would form. For example, Paul McBeth, Ricky Wysocki and Eagle McMahon are three of the many popular faces that make up Cluster 1. These players all ranked in the top 10 during the 2017 World Tour season. Paul McBeth and Ricky Wysocki tied for 1st and Eagle McMahon finished 6th (Spin18, 2017b). The cluster medoid is defined as the observation that has the shortest average distance to all the observations within the cluster (Kaufman & Rousseeuw, 1990). The medoid player would be classified as representative of the elite group during the 2017 World Tour Series. As shown in Table 3, the cluster representative for the elite players was Deven Owens who had a high green hit percentage, inside circle putting percentage, and bullseye percentage, and a low average number of penalties per round.

Rank	GRH.	ICP.	OCPavg	BUE.	PenaltiesPerR
11	47%	83%	0.8	12%	2

Table 3:

Statistics for Deven Owens - Medoid for Cluster 1 (2017), Elite Tier

Clusters 2 and 3 were comprised of less recognizable faces in today's sport. With the challenge of discerning mid tier from low tier players based on name recognition alone, the cluster medoids, or cluster representatives for both clusters were calculated. Jani Kleemola represented Cluster 2. Niklas Eriksson represented Cluster 3. Based off the ranks from these cluster representatives, there seems to be evidence that Jani is in the mid tier (rank 63) and Niklas is in the lower tier (rank 113). As shown in Tables 4 and 5, their relative performance also supports this conclusion. However, the question remains, what factors are leading them to be grouped into their respective tiers?

Rank	GRH.	ICP.	OCPavg	BUE.	PenaltiesPerR
63	30%	79%	1.3	6%	3.14

Table 4:

Statistics for Jani Kleemola - Medoid for Cluster 2 (2017), Mid Tier

Rank	GRH.	ICP.	OCPavg	BUE.	PenaltiesPerR
113	22%	80%	0.3	3%	2.67

Table 5:

Statistics for Niklas Eriksson - Medoid for Cluster 3 (2017), Low Tier

The cluster representatives show two clear differences across their 2017 World Tour statistics. Similar results were observed for the 2016 cluster representatives found in Appendix C. Deven Owens recorded a green hit nearly 50% of time. Jani Kleemola's green hit percentage was much lower at 30%, and Niklas Eriksson's was even lower, residing at 22%. The green, representing a 30-foot radius circle around the basket, is the target for all players when playing a hole. The closer they can land their disc to the basket, the easier it becomes for the player to make their disc in the basket on their next throw. With this relative ordering of the cluster representatives' performance, there is evidence that green hit percentage, predominantly recorded from the initial throw off the tee (drive), is playing a large role in the separation of tiers.

Another clear hierarchy can be seen when observing bullseye percentage. The bullseye is a 5-foot circle located around the pin. When a disc lands in the bullseye, it is almost certain that players will putt their discs in the basket on their next throw; this a crucial factor in lowering one's score. To no surprise, Deven Owens, of the elite tier, had a percentage of bullseye hits that was double the percentage of Jani, the next tier, with Niklas yielding the lowest percentage. This variable is also predominately a measure of how a player is performing off the tee.

Density curves and scatterplots of the data, conditioned on each of the three groups can be seen below in Figure 8. This graphical tool further supports the findings based on the medoids above. Each of the three tiers is represented by their associated color in the scatterplot matrix. The elite tier is blue, the mid tier is green, and low tier yellow. There are three distinct density curves with little overlap for the variable green hit percentage, with the elite tier peaking at higher values, and the low tier peaking at lower values. This same structure can be seen with bullseye percentage, further emphasizing that performance off the tee seems to be a crucial skillset for these professionals when trying to place high. It should be noted that the elite tier also yielded peaks at higher values than the other two tiers for the variables that measure putting performance (ICP. and OCPavg).

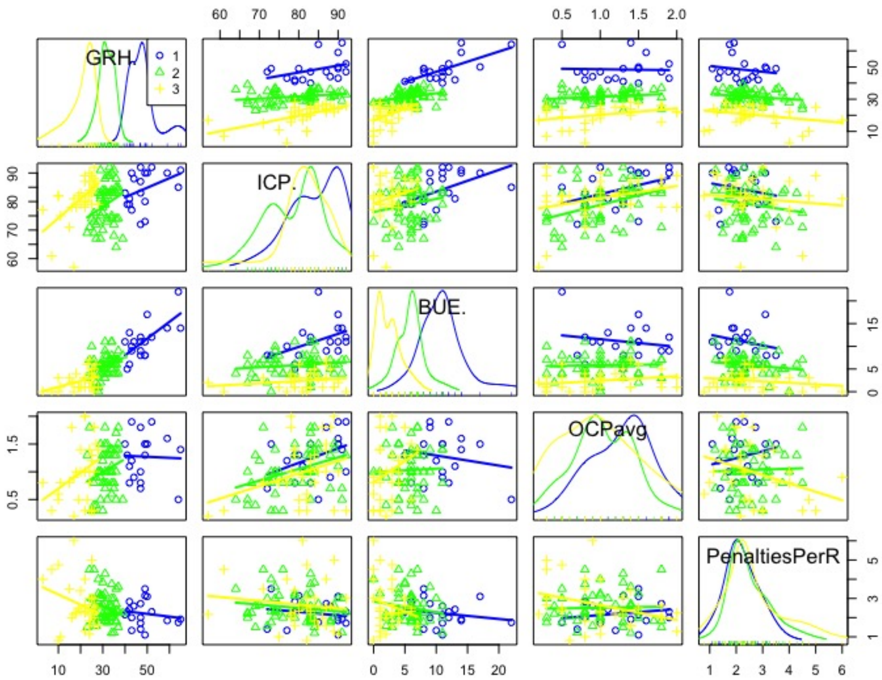


Figure 8: Scatterplot matrix of the variables across the three tiers of players

3.3 Summary

"Drive for show putt for dough" is a famous line coined by Bobby Locke, a four-time open champion in the sport of traditional golf. However, based on the cluster representatives, we have evidence that this may not extend to disc golf. There is no clear hierarchy in the variables associated with how well these representatives putted during the 2017 World Tour. With only four percentage points separating the three tiers' inside circle putting percentages, and the mid tier representative yielding the lowest percentage, it appears that, for these professionals, putting didn't have as much of an impact on tier grouping. Furthermore, the highest score for outside circle putting average, another variable associated with the skillset of putting, was not held by the elite group representative. This provides additional evidence that these tiers, and associated tournament placements seem to be driven by performance off the tee, not putting.

Based on the density curves (Figure 8), the large amount of overlap across the three tiers provides further evidence that the skillset of putting did not play as large of a role in differentiating between these athletes' performance. Similar results can be found for the 2016 World Tour Series, with graphical representations of the data located in Appendix C.

4. Discussion and Conclusions

In major sports competitions, organizations hire analytical teams to help implement different strategies and drive team decisions based on what the data tell them. However, in disc golf, this isn't the case. At least not yet. With the sport growing and prioritizing the collection of data at these World Tour events, there are opportunities for professionals and amateurs alike to gain information on a variety of aspects associated with the sport. This research focused on two main aspects, providing unique perspectives into one's preparation for the game. The data-based tools and perspectives are useful whether you are wanting to develop strategies for an upcoming hole or to study the skillsets of professionals, hoping to gain knowledge and implement what was learned into your own play.

4.1 Study 1: Key Findings, Implications and Future Research

The 3D visualizations provided a different set of information than the still pictures that players currently use when preparing to play a hole. These 3D images could provide players useful information both prior to and during tournament rounds. With the ability to connect characteristics observed while on the course with landscape information being emphasized by the 3D visualizations, players could use this suite of information to help plan each next shot accordingly. Players that developed strategies with the information emphasized from the 3D visualizations were more apt to identify the characteristics of landscape and elevation than when using still pictures. The characteristics noticed inspired different strategies, where players were strategizing around elevation more effectively as a result.

Although the 3D visualizations are useful, they are not intended to replace current preparation strategies. For example, when using the 3D visualizations to develop strategies, players still asked questions about the size of objects, the texture of the landscape, and the scale used. These 3D visualizations could be coupled with other preparation tools in order to explore how to optimize the information a player gains with the fewest amount of resources.

Furthermore, the ability to go mobile with these visualizations could provide players the flexibility to assess landscape, elevation change, and top down views of a given hole while they are on a tee pad waiting to throw. With players often in front of you during a tournament, you have time between holes where you sit and wait on the tee. It is unethical to walk up and down the hole while others are still competing on it, but you could use a mobile 3D visualization to supplement the information you can observe while standing on the tee box.

For practical purposes, the 3D visualizations will be adapted to depict all landscape for a given hole, not just important landscape. In addition, automated imputation of pin placement will be explored so that data collection is no longer a manual process by the creator. This will further expand the number of holes that can be created. This would also create the potential for players around the world to submit data, allowing users to more efficiently expand the number of holes and courses being modeled.

4.2 Study 2: Key Findings, Implications and Future Research

Without many studies on professional disc golf players and the skillsets that separate their performance, this study provides useful information for a variety of players, helping them to identify which skillsets to enhance. However, these results may vary across courses. A course that is more challenging from the tee may lead to results that highlight performance driving off the tee; due to the increased difficulty of the course, this may be the lone separator between players. In contrast, if a course is relatively wide open off the tee, but has challenging putting greens, how well players putt may be the sole separator between players' performance. Therefore, future research should study the holes from which these data were collected to get a sense for why performance off the tee was the main factor separating players' performance in the 2016 and 2017 seasons.

This study uses results from professionals, and makes the assumption that what separates the professionals in rankings can be applied to that of an amateur division. However, this may not always be the case. With little research, and no round-level data to be examined on amateur players, it is hard to justify that such players should focus on one aspect over another. For example, even though all three tiers of professional players had similar putting percentages, there may be differences in putting performance across mixed ability amateurs. In fact, we could be seeing results that indicate regardless of tier, high putting percentages help a player achieve professional status.

References

Becker, B. (2014, January). *Disc golf course review: View and review over 8000 disc golf courses!* www.dgcoursereview.com/media.php?id=4473&mode=media&start=4&page=1&cimg=121450

Danger. (2010, June). *Disc golf course review: View and review over 8000 disc golf courses!* www.dgcoursereview.com/media.php?id=3393&mode=media&start=10&page=1&cimg=47559

Cooley, D. (2018). googleway: Accesses Google maps APIs to retrieve data and plot maps. R package (Version 2.4.0). https://cran.r-project.org/web/packages/googleway/googleway.pdf

Disc Golf Association. (2017). *Disc golf history DGA.* www.discgolf.com/disc-golf-education-development/disc-golf-history

Ersi. (2017). *How kriging works.* pro.arcgis.com/en/pro-app/tool-reference/3d-analyst/how-kriging-works.htm

Everitt, B., & Hothorn, T. (2011). *An introduction to applied multivariate analysis with R.* New York: Springer.

Google Earth Pro ©2018a Google V 7.3.1.4507, (July 25th 2014). South Hills, Helena, Montana.

Google Earth Pro ©2018b Google V 7.3.1.4507, (N/A). Warden Park, Great Falls, Montana.

Google Earth Pro ©2018c Google V 7.3.1.4507, (N/A). Rose Park, Bozeman, Montana.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis.* New York: Wiley.

Professional Disc Golf Association. (2018). www.pdga.com

Soetaert, K. (2017). plot3D: Plotting multi-dimensional data. R package (Version 1.1.1). https://cran.r-project.org/web/packages/plot3D/index.html

Spin18. (2016). *All stats.* Disc Golf World Tour. www.discgolfworldtour.com/statistics/season/2016/all

Spin18. (2017a). *All stats.* Disc Golf World Tour. www.discgolfworldtour.com/statistics/season/2017/all

Spin18. (2017b). *Points.* Disc Golf World Tour. www.discgolfworldtour.com/statistics/season/2017/all

APPENDIX A

Survey Questions:
"Still pictures" was changed to "3D visualization" when players were looking at the 3D visualizations.

1. Based **only on** these still pictures of the hole, what characteristics are you able to consider and what do you notice about these characteristics?
2. What line (trajectory) would you try to throw? What would you try to avoid?
3. Where do you want your disc to make first contact with the ground (after your tee shot)?
4. If you could place your disc 30 feet from the basket, where would you put it? Why?
5. Based **only on** these still pictures of the hole, what lingering questions do you have when preparing to play this hole?

Appendix B

Video Links for 3D Visualizations

Bozeman hole: `https://www.youtube.com/watch?v=NlhCn0Edzgg`

Great Falls hole: `https://www.youtube.com/watch?v=9HAyGXmo42o`

Helena hole: `https://www.youtube.com/watch?v=XbBqa_6M6Uc`

APPENDIX C

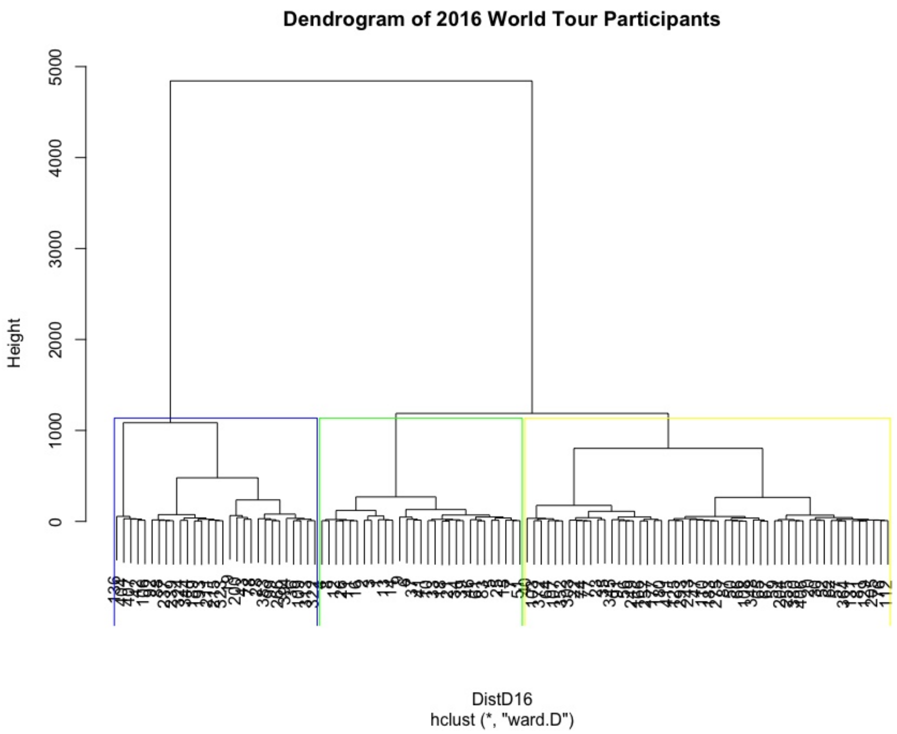


Figure 9: 2016 Disc Golf World Tour Series dendrogram with three clusters

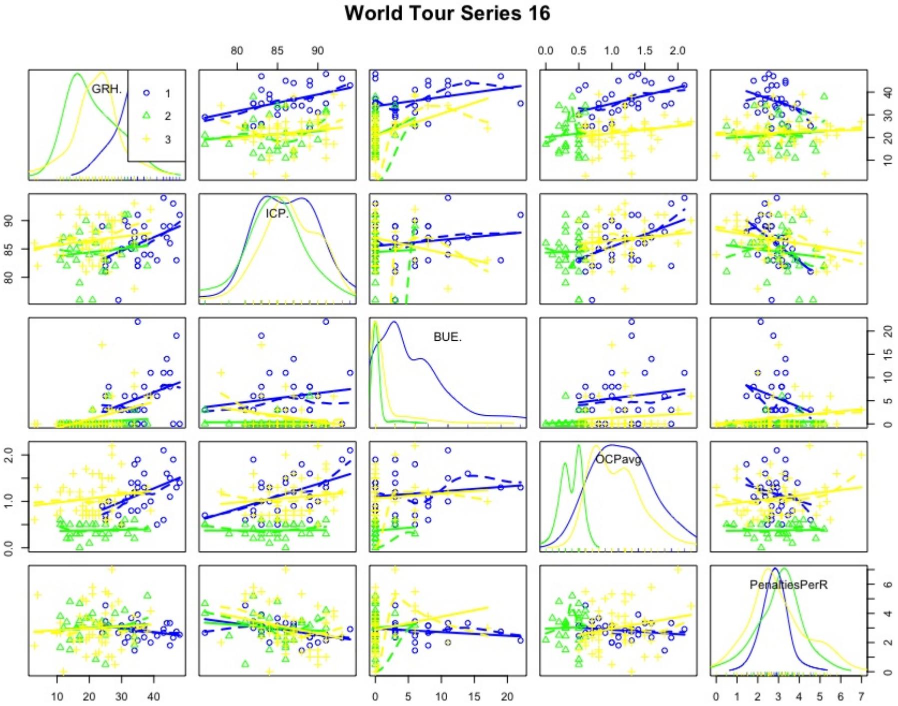


Figure 10: Scatterplot matrix of the variables across three tiers of players

Rank	GRH.	ICP.	OCPavg	BUE.	PenaltiesPerR
35	38%	87%	1.3	6%	2.75

Table 6:

Greg Barsby - Representative for Cluster 1 (2016), Elite Tier

Rank	GRH.	ICP.	OCPavg	BUE.	PenaltiesPerR
71	25%	85%	0.6	0%	2

Table 7:

Teemu Malmelin - Representative for Cluster 2 (2016), Mid Tier

Rank	GRH.	ICP.	OCPavg	BUE.	PenaltiesPerR
112	16%	85%	0.3	0%	3

Table 8:

Arttu Stoor - Representative for Cluster 3 (2016), Low Tier