Golf Performance
Marc Edwards, Yedin Lui, Xinzhe Qi
36-493: Sports Analytics Capstone

The Client
Carnegie Mellon University Golf
Head Coach Daniel Rodgers

Project Goals
With the Carnegie Mellon Men’s and Women’s Golf programs evolving into front-runners in NCAA Division III, Coach Daniel Rodgers is looking to leverage round data collected from each player to identify individual and team strengths and weaknesses.

Data Collection
Each player on the men's and women's team completes a Google Form after each round recording statistics on hole-by-hole score, fairways/greens in regulation, number of putts, approach distance to green, length of first putt, etc.

Data Mutation
The Google Form outputs a spreadsheet that we then load into R. Before performing analysis and creating any visualizations, we modify the format of the raw data to allow for easy manipulation.

Exploratory Data Analysis
After formatting the data, we analyze the information provided through tabulations and visualizations to extract high-level summaries of the data.

R Shiny Application Development
To ensure the data visualization is beneficial to our client, we develop the base code for a user-friendly application mirroring an interactive dashboard.

Application & Data Integration
Upon completion of the shiny app development, we integrate existing analysis and incorporate golf statistics into the application.

Next Season Preparation
Our next steps include establishing standard procedures for data entry, and creating an application that will run with the updated Google Form provided by Coach Rodgers.
Mutating the Data

In our original dataset, each row corresponded to one player who played one round of golf. This meant that all of the data including the holes, pars, scores, approach distance, putts, putt length, and greens and fairways in regulation were all in one row. In order to solve this problem, we utilized a ‘for’ loop to transpose the data and make it easier to read for further use. After creating the new data frame, the data went from 168 observations of 151 variables to 3192 observations of 14 variables. By elongating the data, it made it easier for us to filter and subset the data by using consistent variable names across the 3192 observations.
**Individual and Team Performance Metrics**

**Scoring Average**
Average score across all golf rounds
Goal: 72 strokes per round (regulation par)

**Par Scoring**
Average score recorded across all holes, given par
Goal: Average at or below respective par (3, 4, 5)

**Putting Distribution**
Average number of putts across all holes
Goal: 2 or less putts per hole

**Fairways/Greens in Regulation**
Percent of fairways hit on the drive / Percent of greens hit in regulation
Goal: Greater than 50% of fairways / Greater than 60% of greens

**Strokes Gained: Putting**
Measure of putting performance compared to average PGA Tour player, taking into account length of putt
Goal: Positive strokes gained in putting