2022 Challenge Puzzle

In basketball, the purpose of the game is to score the ball into the basket. During a basketball game, a free throw is awarded to a player when an opponent commits a foul on that player. That player is rewarded with the opportunity to attempt to shoot the ball through the basket without anybody standing in their way. Over the course of an entire basketball game, a player may not get an easier chance to put the ball into the basket than during a free throw.

Yet many professional basketball players – who are supposed to be the best in the world at this sport – struggle to shoot free throws. Some of the most accomplished basketball players in the world have failed to successfully score on free throws even 55% of the time, including Shaquille O'Neal, pictured below. Notice how all his opponents (in white jerseys) are not standing in his way. Still, Shaq struggled to make free throws during his whole career.

Many basketball fans believe that being taller causes players to shoot poorly on free throws. There are multiple theories, such as large hand sizes making it difficult to shoot basketballs.

Your challenge is to construct a data-driven argument for whether height causes poor free throw shooting. In this challenge, you will be asked to answer two questions:

1. **What is the correlation between height and free throw percentage? Please include a data visualization.**
2. **Does height cause poor free throw shooting? Justify your answer with data.**

Data:

You will need to use two data sources:

- A dataset of basketball players' performance, taken from the 2003-2004 National Basketball Association (NBA) season, which you can download here: https://www.dropbox.com/s/u4plikblfy3pwh5/nba.csv?dl=0
- A dataset of basketball players' heights and weights, which you can download here: https://www.dropbox.com/s/jqxyk42el3qbq8b/heights.txt?dl=0

You will find a glossary of variables on the next page.

Assignment:

Please turn in no more than **5 pages (including data visualizations) in PDF format.** You can use any statistical programming software for your analysis, but please include your code as a separate file in a PDF format as well.

You will not need any additional data beyond the two datasets linked above.

Glossary of Variables

nba.csv

Rk	An arbitrary numbering system. Each number is unique to a player.
Player	The name of the basketball player. The name includes a suffix starting with a backslash "\" that can be ignored or removed.
Pos	The position that the basketball player plays. There are five positions in basketball: PG, SG, SF, PF, and C. This Wikipedia article is quick to read and covers the essentials about each position.
Age	The age of the player during the 2003-2004 NBA season.
Tm	The team that the player played for during the 2003-2004 NBA season. Team names are three-letter abbreviations (e.g., Miami is "MIA"). The one exception is "TOT," which is used when a player played for multiple teams during the 2003-2004 season. The "TOT" row is that player's performance for the whole season, whereas the team-specific rows cover that player's performance only with those specific teams.
G	Total games played. There are 82 games in an NBA season.
GS	Total games played as a starter. 5 players start a game for each team (each team may only have 5 players in the game at a time).
MP	Minutes played per game. Each basketball game is 48 minutes.
FG	Field goals made per game. A field goal is a regular shot (includes both 2-point field goals made and 3-point field goals made).
FGA	Field goals attempted per game.
FG%	Field goal percentage. FG% = FG / FGA.
3P	3-point field goals made per game. A 3-point field goal is a subset of FG.
3PA	3-point field goals attempted per game.
3P%	3-point field goal percentage. 3P% = 3P/3PA
2P	2-point field goals made per game. A 2-point field goal is a subset of FG.
2PA	2-point field goals attempted per game.
2P%	2-point field goal percentage. 2P% = 2P/2PA
eFG%	Effective field goal percentage. The formula is (FG + 0.5*3P)/FGA
FT	Free throws made per game.

FTA	Free throws attempted per game
FT%	Free throw percentage. FT% = FT/FTA
ORB	Offensive rebounds per game. Rebounds are awarded when a player collects the ball after another player misses a field goal attempt. ORB refers to rebounds collected while a player is on offense.
DRB	Defensive rebounds per game. DRB refers to rebounds collected while a player is on defense.
TRB	Total rebounds per game. TRB = ORB + DRB
AST	Assists per game. Assists are awarded when a player passes the ball to a teammate who makes a field goal immediately as a result of the pass.
STL	Steals per game. Steals are awarded when a player takes the ball away from the opponent.
BLK	Blocks per game. Blocks are awarded when a player prevents another player from making a field goal by deflecting the ball.
TOV	Turnovers per game. A player commits a turnover when the player loses possession of the ball to the opponent, including as a result of a steal by an opposing player.
PF	Personal fouls per game. A player commits a foul when they do an illegal action. Players are allowed up to 6 fouls per game before they are not allowed to continue playing.
PS/G	Points scored per game. Players score points by making 2-point field goals, 3-point field goals, and free throws.

heights.txt

Name	The name of the basketball player.
Height (Inches)	The basketball player's height, in inches.
Height (Feet-Inches)	The basketball player's height, expressed as feet-inches.
Weight	The basketball player's weight, in pounds.