STAT 331 Midterm Instructions

Policies

- You may use: any item on Canvas (including notes and solutions), any notes you have brought with you, and any resources online.

- You may NOT contact anyone, inside or outside this classroom, during the course of the exam. This includes email, chat/messenger services, and posting on online forums and message boards.
- You may NOT use exam materials from previous years.
- Violations of academic honesty include accessing and using any illegal materials, and giving or receiving help on this exam including looking at other student's exams, allowing other students to look at your own exam, and/or revealing any information about this exam or future exams to someone who has not yet taken it.
- Students who exhibit academic misconduct will be reported to the Office of Student Rights and Responsibilities; academic dishonesty may be punishable by a grade of F in this course.
- You may **not** discuss the exam with any other students until after the exams have been returned to all sections.
- If you witness others exhibiting academic misconduct, you have a duty to report them to your professor.

Instructions

- Please note that you will need to use the data documentation in order to successfully complete this exam.
- The problems on this exam do not necessarily need to be completed sequentially, e.g., if you cannot accomplish problem 1, you may still be able to accomplish problem 2.
- Some parts to problems may need to be completed sequentially. If you cannot complete part (a), please outline code or your thought process for completing part (b).
- All questions should be answered **using R code**. Make sure your code prints out only the information that directly answers the question. **Penalties will be given if your `.html` file has extra output beyond that which answers the question.**
- If your code has errors that you cannot fix in time, you may also include in the comments explanations of your thought process, to potentially recieve partial credit.
- Submit both your `.Rmd` and `.html` files on Canvas by the end of the allotted time. However, only your `.html` file will be graded - make sure the knitted output contains all your answers.
- **Late uploads will be deducted by -10 points per minute!**

Sign below to indicate that you have read and understand these policies and instructions. Unsigned exams will not be graded.

Data Documentation: NBA Player of the Week 1984-2018

NBA facts and vocabulary

In the National Basketball Association (NBA), games are played between Fall and Spring each year, ending with a set of playoff games and a championship. One "season" of games thus spans two calendar years. For example, the 2019-2020 season began on October 22, 2019 and will end on April 15, 2020. The playoffs will begin on April 18, 2020, and will end with the NBA Finals in June 2020.

Teams in the NBA are organized into two Conferences: East and West. From time to time, as teams change locations or new teams are added, the conferences change.

Players join the NBA by being *drafted*, typically when they are 18 years old.

This dataset

The dataset in this exam contains information about the "Player of the Week", awarded each week to the player(s) who performed the best during that week's games.

Each observation in this dataset represents an instance of a certain player being awarded Player of the Week.

The variables in this dataset are:

NBA_player_of_the_week.csv

Age	Player age at the time
Conference	East/West/Missing
Date	Award date
Draft Year	Player draft year
Height	Height (feet)
Player	Full name
Position	Original position
Season	Full season
Season short	Season ending year
Seasons in league	Seasons in league up to date
Team	Team
Weight	Weight (pounds)
Real_value	If two awards given at the same week (East & West) the player got 0.5, else 1 point.

Part One: Data Cleaning and Adjusting

Use the `nba_player_of_the_week.csv` dataset for this section.

Write all answers in `Stat_331-Midterm-Part_One.Rmd`. Turn in both this file AND the knitted `.html` output.

1) Note that the variable `Height` is inconsistent; that the last 100 or so rows of the dataset have height values in centimeters, and the rest in feet and inches. We will fix this first.
(1 foot = 12 inches, 1 cm = 0.3937 inches)

   - *(20pts)* Write a function called `fix_height` that takes as input a single height, in the form of a string. Your function should return a numeric value of height in inches, rounded to the nearest one decimal. For example, inputting the string "5-9" should return the number `69`, and "221cm" should return `87`.
   - *(10pts)* Update the variable `Height` so that all player's heights are given in inches, using either a `map` function or an `apply` function. (You may use a `for` loop for a -4 point penalty.)

2) *(20pts)* Now, note that the variable `Weight` has a similar issue as `Height`: some weights are listed in kilograms rather than pounds. *(1 kilogram = 2.20462 pounds)* Fix this, so that when you are finished, the variable `Weight` will contain only player weights in pounds. You may use any process you wish to accomplish this; however, +2 points of extra credit will be awarded to correct answers that use a vectorized function.

Part Two: Short Answer

The `nba_cleaned` dataset has the heights and weights already cleaned. You may use this dataset to verify your answers to the above. You may also rely on this dataset for the remainder of the exam, if you are not confident of your answers in Part One.

Use code to answer the following questions. For variables that have missing values, you may drop the missing values in your calculations.

Submit all answers to the quiz on Canvas. You may resubmit the quiz as many times as you like, and your scores will be averaged. (You may not, however, resubmit the quiz once you reach 100%)

Be sure to also upload your .html and .Rmd to Canvas, as verification.

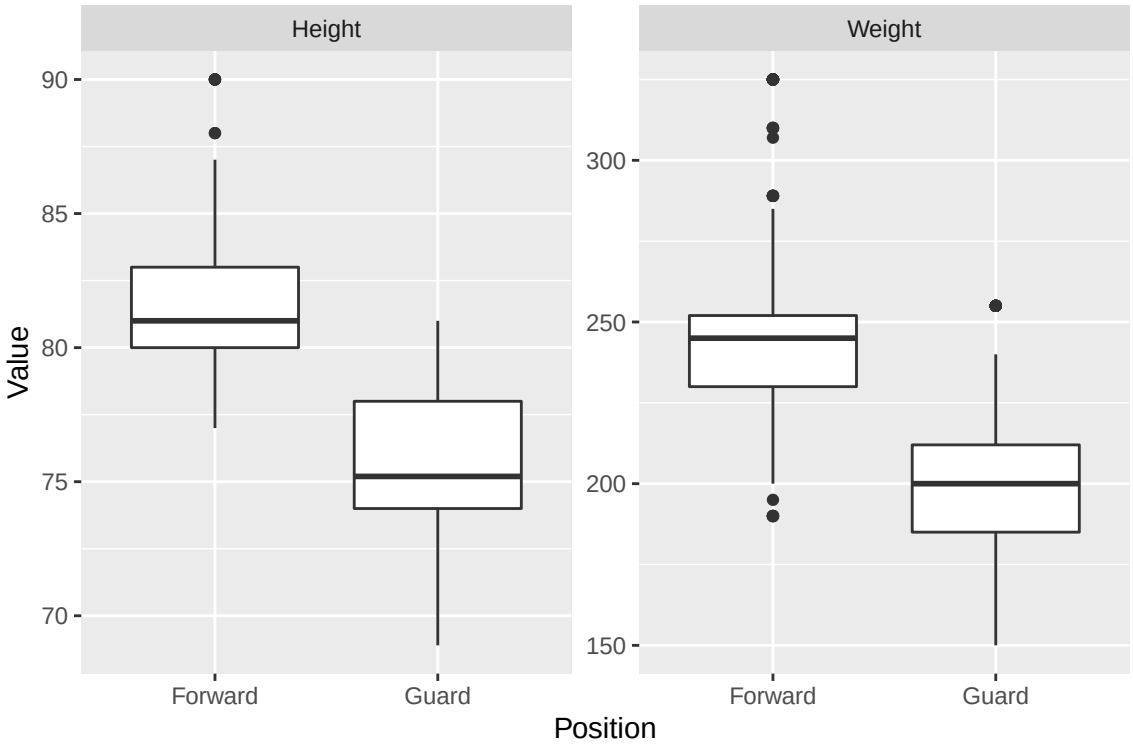
1) *(5pts)* How many rows are in this dataset?
2) *(5pts)* How many unique players won "Player of the Week" in the timespan covered by this dataset?
3) *(5pts)* What are the maximum and minimum ages of players in this dataset?
4) *(5pts)* How many times were two "Player of the Week" awards given in the same week?
5) *(5pts)* Which team has had a player win the award the most times?
6) *(5pts)* Which player has won Player of the Week the third most number of times, after LeBron James and Kobe Bryant?
7) *(10pts)* What is the mean height (in inches) of all players who appear in this dataset?
8) *(10pts)* What is the difference in median weight of players in the Eastern and Western conferences who appear in this dataset?

Part Three: Data Analysis

Write all answers in `Stat_331-Midterm-Part_Three.Rmd`. Turn in both this file AND the knitted `.html` output.

1) Consider the "guard" positions (PG, SG, G) compared to the "forward" positions (F, C, F-C, FC, G-F, GF, PF, SF). Suppose we wish to know if players at these positions tend to be different sizes.

   - *(10pts)* Create a new variable that labels whether the player is a guard or a forward.
   - *(20pts)* Has there been a difference in the importance of Guards vs Forwards over time? Make a plot that explores the POW wins from these positions over the years.
   - *(20pts)* Recreate the following plot.



2) *(20pts)* Define a "superstar" to be a player that wins Player of the Week more than 10 times. Which two teams have the most Player of the Week wins from players who are NOT superstars?

3) Let's consider the 2017-2018 NBA season only. We would like to determine if there were teams that performed better in the second half of the season than the first half.

   - *(10pts)* Find the date that was the exact halfway point of the season; that is, half the games were played before it, and half after it. *If you are unable to solve this question, use* `January 1, 2018` *as your halfway point for parts (b) and (c).*
   - *(20pts)* Which teams had more wins in the second half of the season than in the first?

Extra Credit

Every year, an "All-Star" game is played in February. The dataset `nba_all_star.csv` contains a list of the dates of these games since 1984, and the MVP(s) of each game. Were any players ever the MVP of the All-Star game *and* the NBA Player of the Week the next week?