

# Week 13 Submission

Daniel Tan

2023-11-24

## Final Submission

(1) What is the theme of your data story?

The theme of my data story was to question the commonly held stereotype that the taller one is, the more likely he/she is to do well in Basketball

(2) Why is it important to address this question?

In the immortal words of rapper Skee-Lo, “I wish I was a little bit taller, I wish I was a baller”. Many men worldwide suffer from ‘Napoleon Complex’. A theory suggesting that individuals of shorter stature, particularly men, may overcompensate for their height by displaying overly assertive or aggressive behavior. In a study conducted on the relationship between satisfaction towards one’s height and traits like “psychopathy, narcissism and Machiavellianism”(Kozłowska, 2023), “all three traits were associated with the wish to be taller and shorter stature”(Kozłowska, 2023).

Essentially, the jealousy and inferiority complex shorter men might hold over their height leads to them behaving aggressively and negatively to compensate for it.

By analysing the relationship between height and performance in the NBA, I wanted to show that height wasn’t the end all be all for someone. Even if one’s shorter stature might hold them back in physical activities, successful players pivot and play to their strengths, and are still able to make significant contributions to their team.

(3) Why do you think the data sources that you have curated can help you answer the question?

The dataset(Cirtautas, 2023) I curated, includes a wide span of player data all the way from 1996 to 2023, with such diverse and detailed data, I was able to better understand why some players were successful beyond just the points they scored on average.

(4) What are the insights from the data and how are they depicted in plots?

Contrary to conventional wisdom, which often associates success in basketball with towering stature, the findings suggest a nuanced relationship between height and performance.

The linear regression analysis, examining the effect of player height on net rating, revealed a minimal and statistically insignificant impact.

The scatter plot visually reinforced the diverse performance levels across players of varying heights. Notably, exceptional players like Muggsy Bogues and Earl Boykins, both significantly below the average NBA height, demonstrated outstanding contributions to their teams, challenging the stereotype that height is a prerequisite for success.

The plot showed there wasn't any particular relation between height and performance, with good and bad players all across the board. In fact, a good chunk of the best performing players happened to be under the median height of 200.66cm.

Possible explanations for these results include considerations of injury susceptibility among taller players or challenges in speed and agility. Additionally, the analysis hinted at potential biases in recruitment practices, where shorter players, despite their exceptional skills, may face greater scrutiny or be overlooked in favor of taller prospects.

In conclusion, while height may play a role in certain aspects of the game, the evidence presented suggests that it does not serve as a decisive factor in determining a player's overall impact, particularly when assessed through the lens of net rating. Exceptional skills, teamwork, and individual contributions, regardless of height, continue to shape success in the dynamic and multifaceted world of the NBA.

- (5) How did you implement this entire project? Were there any new concepts that you learnt to implement some aspects of it?

I implemented this entire project using the various methods of data manipulating and creating visualizations taught throughout the semester. I will say that chatGPT was incredibly helpful in answering queries and debugging issues that were otherwise annoying to google for.

Here were the things I struggled with the most so you might address it in future semesters.

File naming conventions and file locations. I got stuck on the very first week of working on this project because I just couldn't for the life of me get GitHub to create the webpage. I am specifically referring to how I followed the step by step tutorial given in Canvas but still faced difficulties. I had all my NM2207 materials in a folder on my desktop. I followed the instructions of "Create\_Publish.pdf", I named my folder accordingly. And at first I thought I had solved the issue by making my repo public(it was private), but GitHub still failed to deploy my webpage. Basically, I had a bunch of folders all named "idrinktapwater.github.io" and that confused GitHub. I had to go and cut and paste all my materials from my desktop folder into the GitHub folder before deleting the desktop folder, ensuring that there was only one folder named "idrinktapwater.github.io" which existed in my GitHub folder in Documents.

Even after that, GitHub pages still failed to publish my webpage. I figured out that I could check the debugging logs, and found that for some reason GitHub was trying to render my folder named "week1to13" along with the quarto documents. Putting an underscore at the front of the folder/file name fixed this issue, but how was I supposed to know that?

Also I struggled with finding a good topic to make a data story about, with the caveat of "why is this question important" paralysing my momentum. Because personally the overlap between what I found interesting and worth exploring and what might be conventionally recognised as "important", was very little.

My next biggest struggle was trying to get my app to work on the Webpage. I didn't realise that just because something is able to run in R when you have access to a global environment, that it won't work once you try to render the individual document! I eventually figured out you had to recreate the steps you took in creating whatever variables you called on in the code for the render to work properly.

```
# Create a data frame (table)
my_table <- data.frame(
  Weeks = c("2 and 4", 3, 5, 6, 7, 8, "Self-taught", "Self-taught", "Self-taught", "Self-taught", "Self-taught"),
  Topics = c("Dataset Manipulation", "Variables", "Functions", "Iterations", "Data visualization", "Shiny")
)

# View the created table
print(my_table)
```

```
##           Weeks
```

References: Cirtautas, Justinas. NBA Players. <https://www.kaggle.com/datasets/justinas/nba-players-data>. Accessed 24 Nov. 2023. Kozłowska, Monika A., et al. ‘The Napoleon Complex, Revisited: Those High on the Dark Triad Traits Are Dissatisfied with Their Height and Are Short’. *Personality and Individual Differences*, vol. 203, Mar. 2023, p. 111990. ScienceDirect, <https://doi.org/10.1016/j.paid.2022.111990>.

## Week 9

<https://www.kaggle.com/datasets/yakhyojon/national-basketball-association-nba>

3

How much does height matter at the highest level of play?

##(2) Why is this an important question? (Answer: 3 sentences, each of which has some evidence, e.g., “According to the United Nations...” to justify why the question you have chosen is important),

First, historical data and analysis have consistently shown that height plays a pivotal role in a player’s success in the NBA. Taller players tend to have advantages in areas like rebounding, shot-blocking, and overall defensive performance. This trend is supported by statistics, which reveal that the majority of NBA MVPs and All-Star selections over the years have been above-average in height.

Second, understanding the significance of height in the NBA has broader implications for issues like talent development, scouting, and player recruitment. NBA teams invest substantial resources in selecting and nurturing talent, and recognizing the impact of height can lead to more informed decisions about which players to draft and develop.

Lastly, the question touches on broader themes of fairness and inclusivity in sports. By investigating the role of height, we can also explore questions related to diversity, as shorter players who defy height-based expectations provide a compelling narrative in the league. Analyzing this question is not just about basketball; it’s about evaluating societal perceptions and norms that can extend beyond the court.

##(3) Which rows and columns of the dataset will be used to answer this question? (Answer: Actual names of the variables in the dataset that you plan to use).

player\_name, player\_height, pts, reb, ast, ast\_pct

## Week 11

- (1) List the visualizations that you are going to use in your project (Answer: What are the variables that you are going to plot? How will it answer your larger question?)

First I’ll create a new column “Avg NBA player” whose row will just consist of the mean for each column e.g Height, Pts, Assists etc. Then I’ll plot the statistics for the players shorter than the average height against this theoretical “average NBA player”.

(6/11/2023) I plan on adding colour grading into my charts so it’s easy to see at a glance what stats are considered good/bad for the layman without intimate knowledge of the sport.

- (2) How do you plan to make it interactive? (Answer: features of ggplot2/shiny/markdown do you plan to use to make the story interactive)

I plan to have an interact-able on the website similar to what is found on “mygamatoto.com” that app essentially just lets you pull up the row of any player the user wants to select, that can mainly be used for eyeball comparisons between different players the user chooses to inspect.

It’d be fun to have a guessing/trivia game as well, where I can ask questions about certain players, and the user has to guess a value like “The ‘average’ NBA player scored X points in Y season, how many do you think Z player scored that season?”

- (3) What concepts incorporated in your project were taught in the course and which ones were self-learnt? (Answer: Create a table with topics in one column and Weeks in the other to indicate which concept taught in which week is being used. Leave the entry of the Week column empty for self-learnt concepts)

```
# Create a data frame (table)
my_table <- data.frame(
  Weeks = c("2 and 4", 3, 5, 6, 7, 8, "Self-taught", "Self-taught", "Self-taught", "Self-taught"),
  Topics = c("Dataset Manipulation", "Variables", "Functions", "Iterations", "Data visualization", "Shiny")
)
```

```
)
```

```
# View the created table
```

```
print(my_table)
```

##	Weeks	Topics
## 1	2 and 4	Dataset Manipulation
## 2	3	Variables
## 3	5	Functions
## 4	6	Iterations
## 5	7	Data visualization
## 6	8	Shiny
## 7	Self-taught	Good Folder naming conventions
## 8	Self-taught	write.csv (how to save a csv that I've manipulated)
## 9	Self-taught	purrr walk2
## 10	Self-taught	filter(net_rating == max(net_rating))

Include the challenges and errors that you faced and how you overcame them.