Final Submission

Yang Juan Hun

2023-11-19

Week 9

The topic I have finalized on is Dungeons and Dragons Fifth Edition (DnD 5e), a tabletop role-playing game (TTRPG). The specific data I am looking at is character creation choices from approximately 8,000 DnD 5e players all over the world.

The data sources I will be using are from https://github.com/oganm/dnddata, a Github repository which compiles data on character choices in DnD 5e into an R package.

Week 10

The question I intend to answer is "What are the average ability scores, hit points, and armor classes of certain DnD 5e character builds at specific levels?"

This is an important question, because according to Reddit, players who are new to DnD 5e may have trouble creating their character and could benefit from knowing what general trends are like for most other players. Additionally, according to the DnD 5e Player's Handbook, the game developers recommend that the ability scores of a character should be used to guide how a player roleplays that character, thus having more information on how most people allocate their ability scores for specific character builds can guide how a player conceptualizes the personality of their character. According to enworld.org, some TTRPG players also enjoy planning for future levels when creating their characters, thus knowing general statistical trends of higher level characters can help such players with their pre-planning and optimizing.

> Sources:
>
> - https://www.reddit.com/r/DungeonsAndDragons/comments/14if2oe/are_these_stats_overpowered_im_new_to_dnd_and_my/
> - https://dnd.wizards.com/products/rpg_playershandbook
> - https://www.enworld.org/threads/character-progression-planning-how-do-you-do-it.668394/

The columns I intend to use are "level", "justClass", "processedRace", "HP", "AC", "Str", "Dex", "Con", "Int", "Wis", and "Cha". The rows I intend to use are all rows, with 'NA' values and significant outliers omitted, such as row 88 who has unrealistically high statistics such as over a thousand hit points.

One challenge I faced while compiling this data is that the "justClass" column can contain multiple values at once. In DnD 5e, there is a mechanic known as 'multiclassing' that allows a player's character to have multiple classes at once. Some examples of DnD 5e classes include "Artificer", "Fighter" and "Wizard". There are examples of players in the database who have multiclassed into all three aforementioned classes, resulting in their value under "justClass" displaying as "Artificer|Fighter|Wizard". To rectify this challenge, I mutated the dataset to include new individual columns for each class in DnD 5e, the variable names being

the class names. The values found under these columns would be either "TRUE" or "FALSE", indicating whether or not the player's character had that specified class.

Another challenge I faced while compiling this data was, unsurprisingly, NA values and outliers. One clear example is row 88. The statistics of that character are comedically high compared to all other entries in the database, so I had to carefully sift out such outlier characters from the database.

Week 11

The visualization I am intending to use is primarily a bar graph depicting the frequency of a certain DnD statistic (here, the term 'DnD statistic' refers to HP, AC, Str, Dex, Con, Int, Wis and Cha), given a specific class (or classes) and range of levels. This allows a user to easily visualize highly customizable data, providing a wide range of insights into how DnD statistics are distributed across a large sample size, depending on the characteristics of the player's characters. For example, a player who wishes to get a grasp of the average Str scores of Fighters between levels 1 - 4 (those levels are often referred to as 'Tier 1' gameplay) can easily do so with this app by inputting those exact specifications into the interface. Thus, it aids in answering the overarching question of "What are the average DnD 5e statistics of certain character builds at certain levels".

> Source: https://arcaneeye.com/dm-tools-5e/what-are-the-tiers-of-play-in-dd/

I intend to make the app interactive through the highly customizable interface, using Shiny. I am using multiple types of inputs, such as group checkbox inputs, radio buttons and sliders to give the user highly intuitive control over the visualization of the data. The data itself will be visualized using a bar graph generated with ggplot2. Using these inputs, the user will be able to choose what DnD statistic they are analysing, what classes they would like to analyse, whether they will include characters with at least one of the selected classes or only the selected classes, the minimum and maximum level of the characters they are analysing, how sensitive the outlier detection will be (the scaling factor of the IQR), and the type of graph they wish to view (bar graph or histogram). If they wish to view a histogram, they can also customize the number of bins shown.

Topics	Week
Filter	1
Select	3
Summary statistics	4
Mutate	4
Pipeline operator	4
Custom functions	5
For loops	6
ggplot2	7
Shiny App	8
Converting an untidy multiple-choice categorical variable into a logical matrix	
Identifying outliers	
Identifying NA values row-wise	
grepl()	
Observation events in Shiny	
Updating Shiny Inputs reactively	
Creating Conditional Shiny Inputs	
Complex logical filters e.g distinguishing between any of the selected, only either or all selected, and specifically all selected	
Interactive plots that can react to mouse clicks	
How to calculate mode manually	

Topics	Week
How to plot bar graphs with two variables	

Previously, I intended to simply remove outlier data, however I came to the realization that due to the nature of DnD 5e, extreme variance in the overall dataset without any additional filters is to be expected (gameplay at lower levels is vastly different to gameplay at higher levels), further more there was a much higher concentration of low-level players in the dataset than high-level players, skewing the data. Thus, it would be unwise to simply remove outliers from the entire dataset as that would remove genuinely useful information on high-level play. To solve this, I instead opted to have the outliers be removed only when printing the bar graph, instead of removing the outliers from the original dataset. In addition, I expanded this into an additional control for the user, allowing them to customize how sensitive outlier detection can be for them by adjusting the scaling factor of the IQR (Outliers are calculated as $>$ 3rd quartile + scaling factor * IQR or $<$ 1st quartile - scaling factor * IQR. Scaling factor is set to 3 by default).

Another challenge I faced was the creation of a 'minimum-maximum' system. If a user set the minimum level to be higher than the maximum level, that woul result in the app not providing any visualization. To rectify this, I added a system that would update the maximum value of the minimum level slider and the minimum value of the maximum level slider, such that a user cannot exceed said limits within the system to produce nonsensical inputs.

Week 12

As I continued to improve on my website's design and the interactive app, I discovered some flaws with my original iteration of my data visualizer.

For one, I realised my prior method of dynamically changing the maximum and minimum values of the minimum/maximum level sliders was visually very confusing and disorienting, as the length of the sliders would not change even though their ranges would. I instead tweaked my approach, such that whenever the minimum level slider was set to a value above the maximum level, it would snap its value to the maximum level instead, and vice versa for the maximum level slider.

Another flaw was that I showed the details of observations that were identified as outliers, but I did not have any such section for observations that were not outliers. To accommodate this, I added a segment for showing the details of observations in the dataset, along with implementing a 'page' system for determining how many observations you can view in one 'page', from which you can flip through the 'pages' to view all the observations in the filtered dataset.

I also discovered that my initial conceptualization of the class selection system was slightly flawed - I original thought I could give users the option to view characters with 'at least one of the selected classes' or 'with only the selected classes'. As it turned out, my method of filtering for characters with 'only the selected classes' was actually a method of filtering for characters with 'exactly the selected classes'. For instance, if I selected the 'Fighter' and 'Barbarian' classes and chose to view 'only the selected classes', I would only be able to view characters who were multiclassing into both 'Fighter' and 'Barbarian', and it would not show any pure Fighters or Barbarians. To rectify this, I renamed that original 'only' option to a more fitting 'exactly the selected classes' option, and added a new 'only' option that would logically filter the dataset such that if I selected the 'Fighter' and 'Barbarian' classes, I would yield all Fighter-Barbarian multiclassing characters, as well as pure Fighters and Barbarians as well.

I also realized that it was sometimes not very easy to interpret the graph visually. To rectify this, I added a system where you could click on a bar in the graph, and the app would print out the exact values of the X variable (chosen DnD stat) and Y variable (count) of that bar at the bottom of the graph for ease of interpretation.

Beyond rectifying those challenges in my prior data visualizer, I also began work on a secondary data visualizer, intended to help highlight the insights that can be derived from my first data visualizer and

depicting its findings and potential conclusions in an interactive format. This second data visualizer would look only at the 'Fighter' class, as it was the most common class in the dataset, and it would depict how the means, medians or modes of the six core ability scores (Strength, Dexterity, Constitution, Intelligence, Wisdom and Charisma) of Fighter characters at specific tiers of gameplay (levels 1-4, 5-10, 11-16, 17-20) compare to each other side-by-side on a bar graph. This would help show insights into how most fighter players distribute their ability scores at specific tiers of gameplay, which could aid a new player intending to play a fighter on how they could distribute their ability scores.

> Source: https://arcaneeye.com/dm-tools-5e/what-are-the-tiers-of-play-in-dd/

This secondary data visualizer is not yet in a functional form and thus has not been added to the website.

In addition, I learnt new concepts in R and Shiny that weren't covered in lecture. These new concepts have been added to my prior concepts table from week 11.

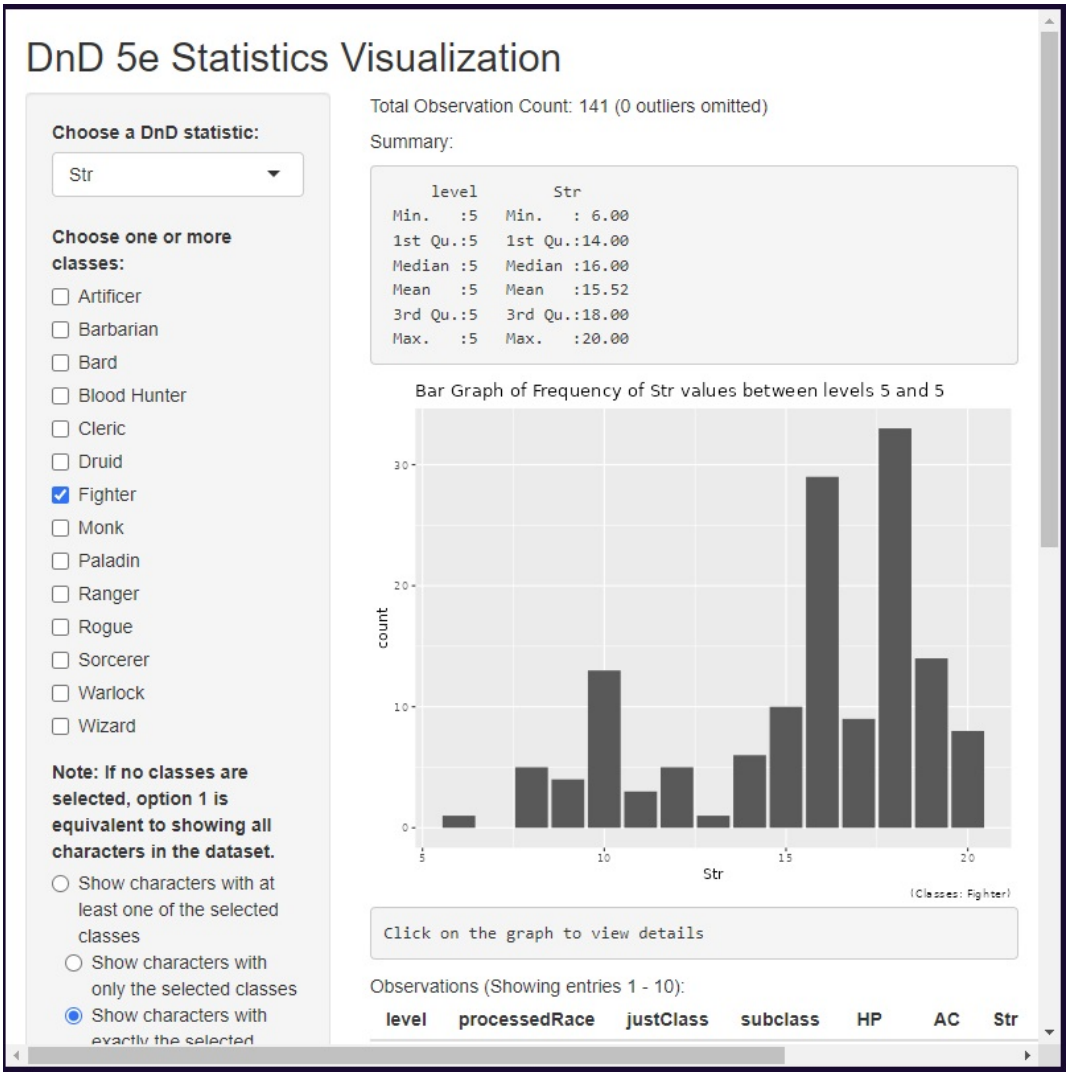
Final Submission

The theme of my data story is Dungeons and Dragons Fifth Edition, a tabletop role-playing game. The question I intend to answer is "What are the average ability scores, hit points, and armor classes of certain DnD 5e character builds at specific levels?"

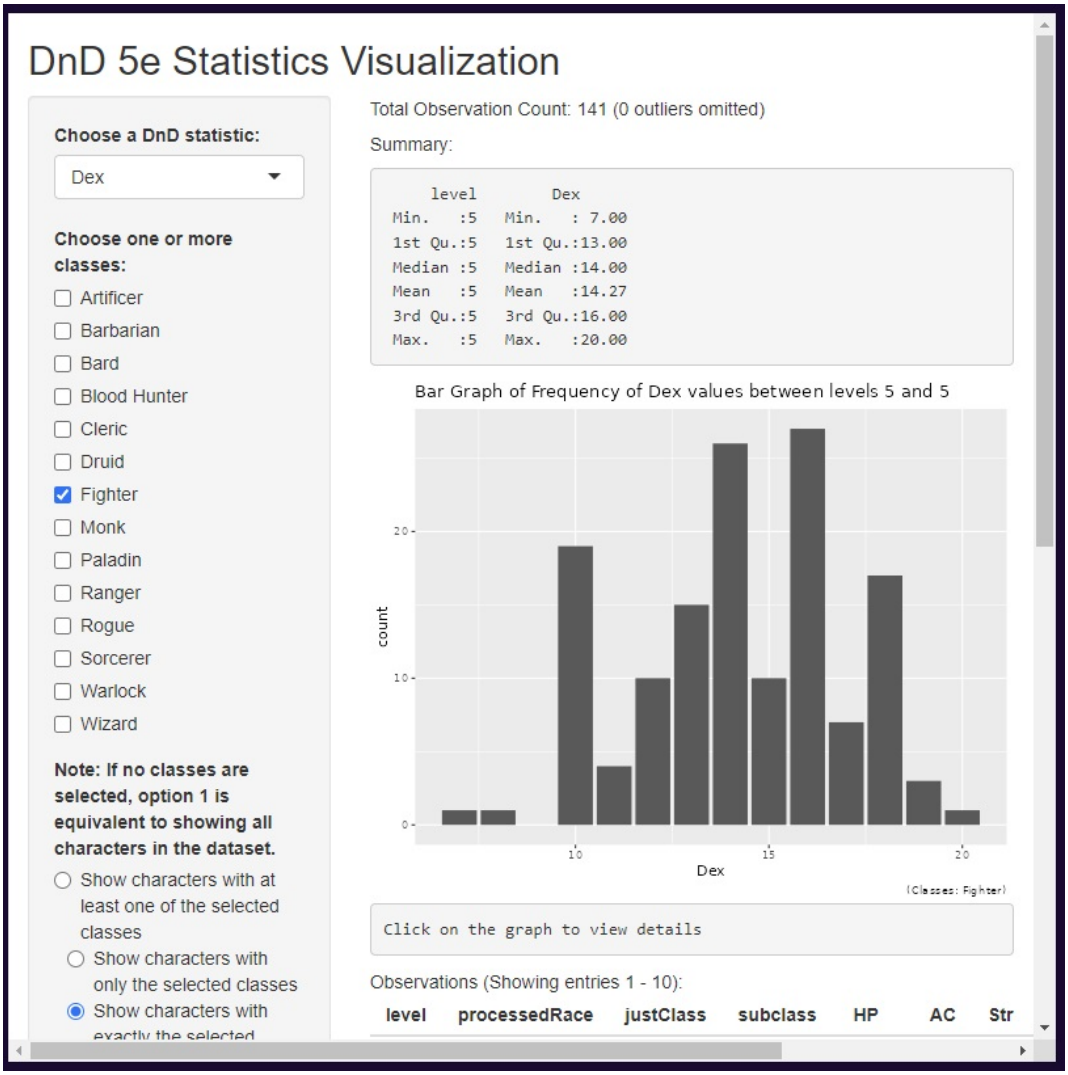
This is an important question, because new DnD 5e players may have trouble creating their character and could benefit from knowing the general trends of other player characters ("Are these stats overpowered?", 2023). Additionally, the game developers recommend that the ability scores of a character should be used to guide how a player roleplays that character, thus knowing how most people allocate ability scores for specific character builds can guide how a player conceptualizes their character's personality (Wizards of the Coast, 2023b). Some TTRPG players also enjoy planning for future levels during character creation, thus knowing general trends of higher level characters can help such players with their pre-planning and optimizing ("Character progression planning", 2019).

The data source I used was from a Github repository which compiles data on character choices in DnD 5e into an R package (Oganm, 2022). The specific dataset I used compiled character choices from approximately 8,000 DnD 5e players all over the world (a total of 53 different countries). As it was a fairly large sample size with a diverse spread of subjects, it should provide adequate insights into character creation trends for a wide variety of builds and levels.

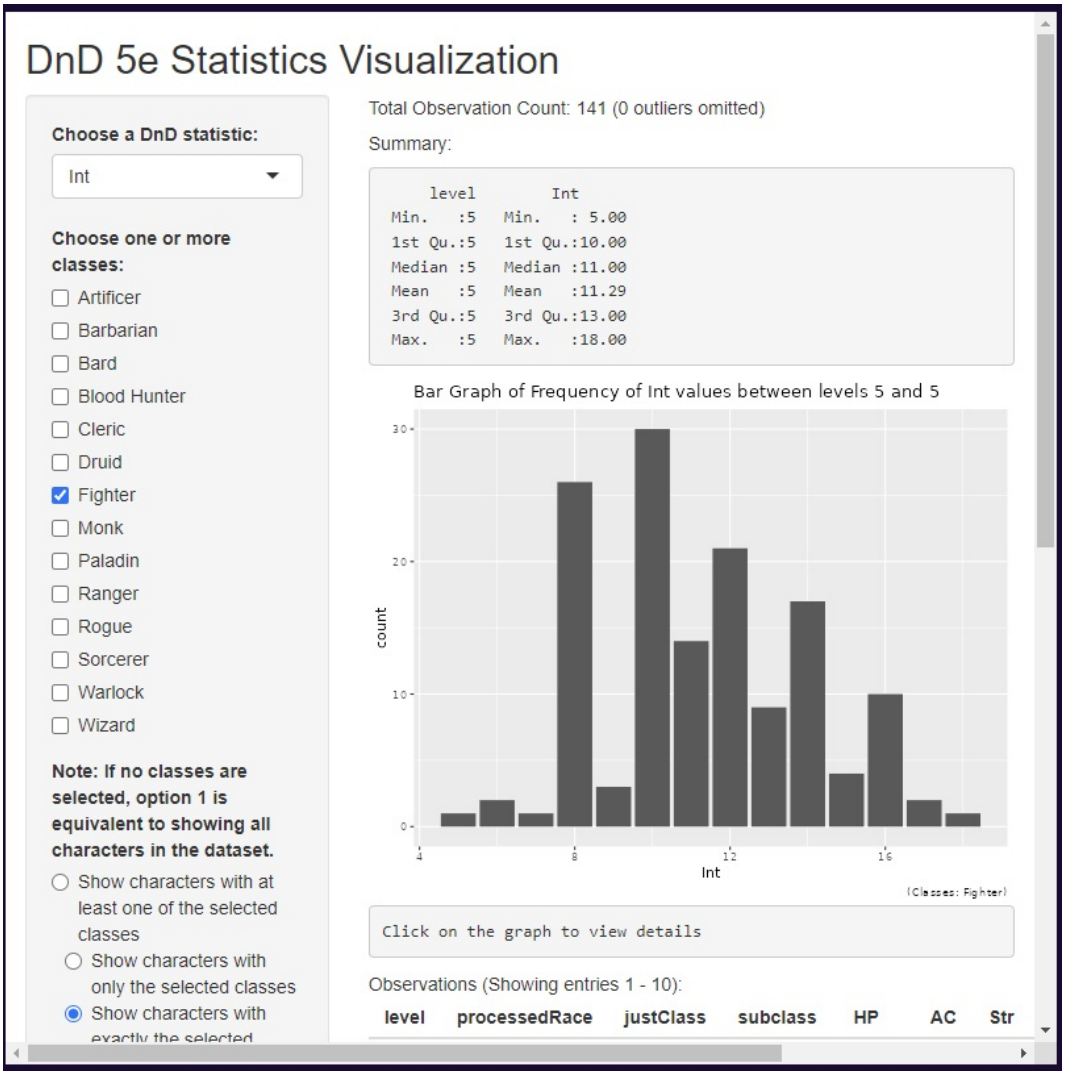
When focusing on the Fighter class, many insights can be derived. For instance, Fighters tend to be biased towards high Strength scores. This is depicted in bar graphs of Fighter Strength scores against those scores' frequencies in the dataset, where we can see that such a graph is skewed to the right. Note that the bar graph controls for level where we are only looking at level 5 Fighters, as the ability scores of DnD characters at different levels will vary vastly due to resources unlocked at later levels (Wizards of the Coast, 2023b).



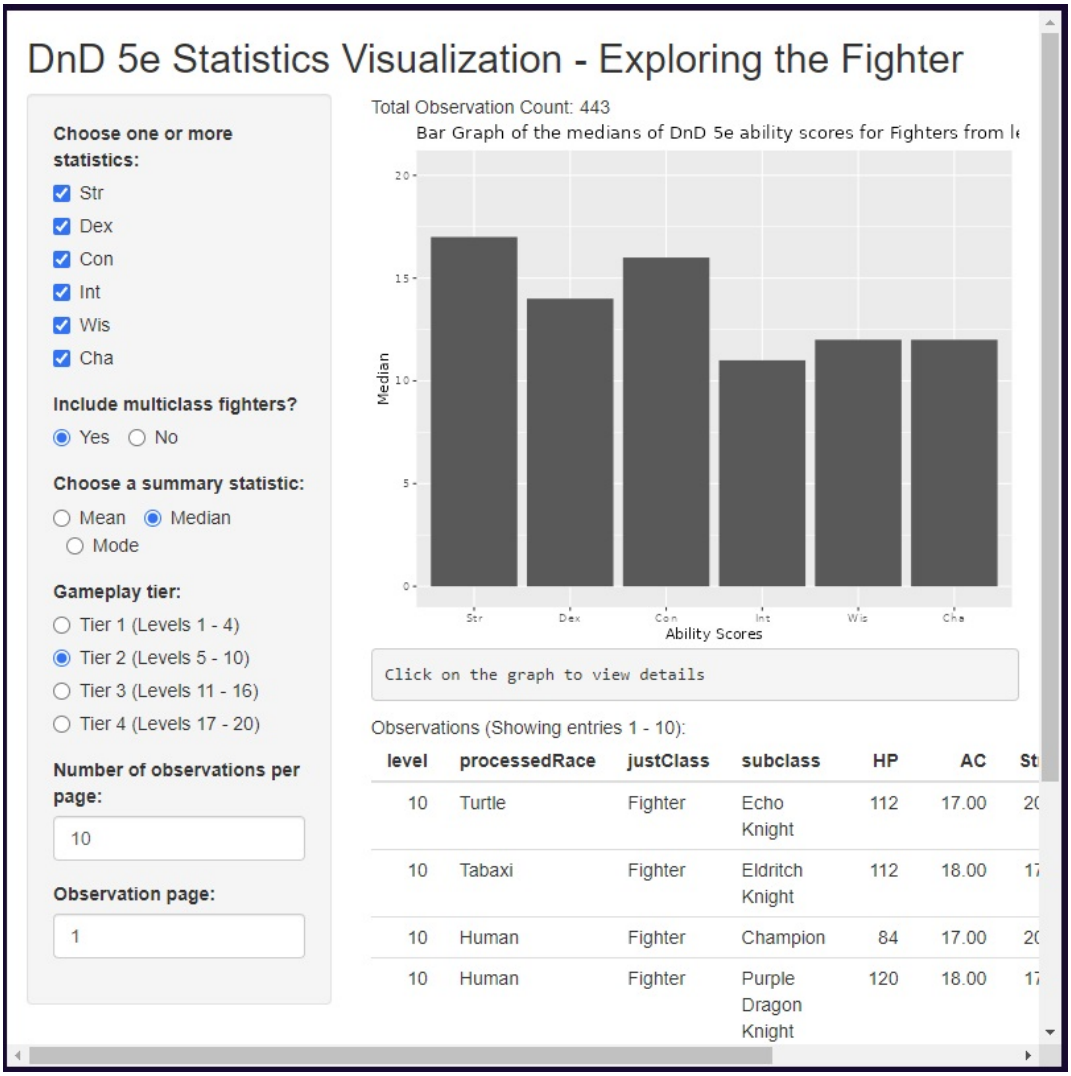
When plotting Dexterity scores in such a bar graph, we find that Dexterity scores have multiple notable peaks in their distribution. This suggests that there are perhaps multiple different play-styles for Fighters with varying focus on Dexterity.



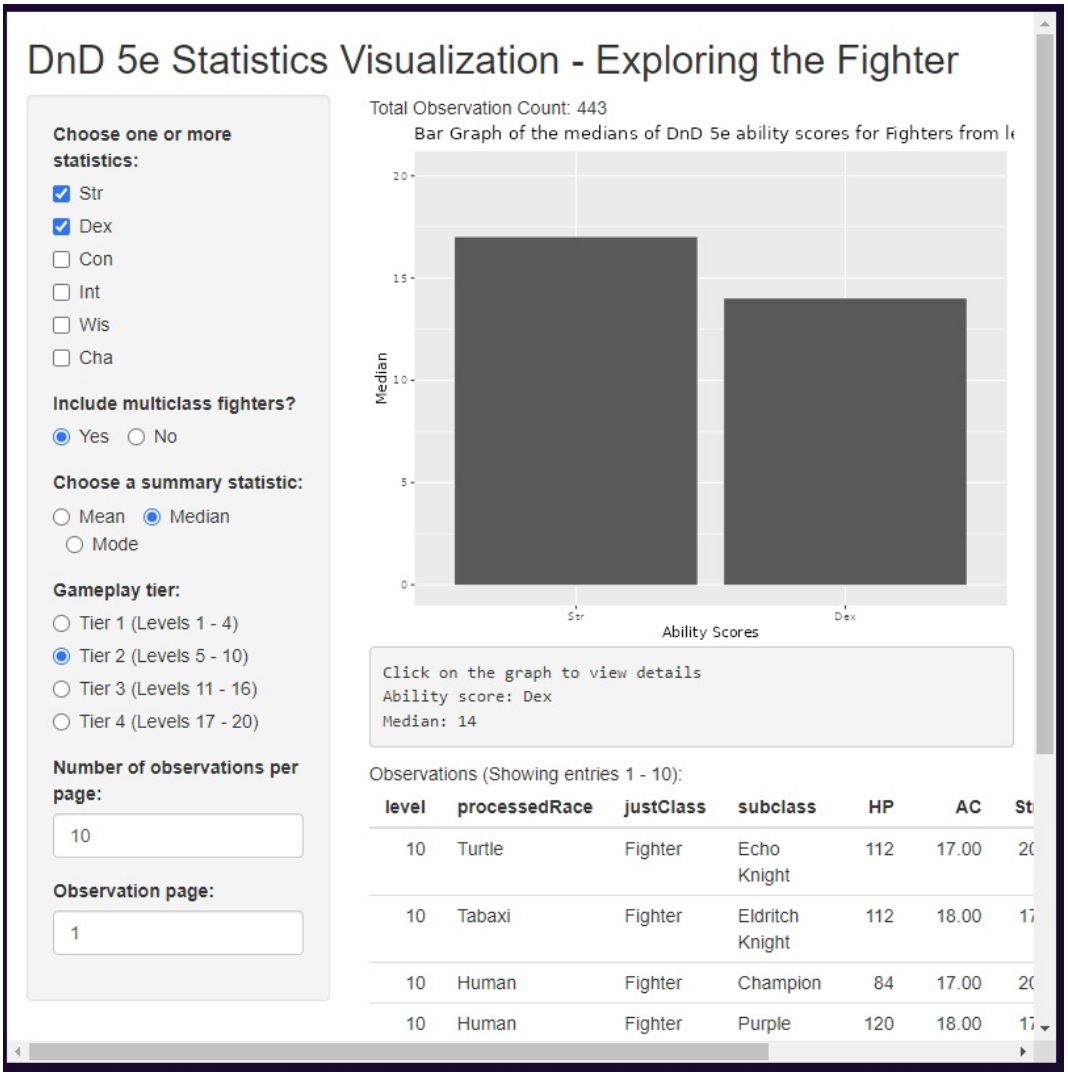
Intelligence scores, meanwhile, seem slightly skewed to the left when plotted in this bar graph. This might suggest that Fighters tend towards lower Intelligence scores.



We can derive further insights using another bar graph that plots the medians of the six core ability scores of Fighters side-by-side, controlling for gameplay tier at Tier 2 (i.e. only analyzing levels 5 - 10). This revealed that Fighters invest more into their Strength, Dexterity and Constitution than their Intelligence, Wisdom and Charisma, which reflects ideal ability score distributions for such a combat-focused class.



A more interesting insight is how Strength compares to Dexterity. Strength is typically more invested in for close-range combat character builds, whereas Dexterity is more invested in for long-distance combat character builds. These are two play-styles that Fighters can be equally effective in. The data however suggests that Fighter players are biased towards close-quarters combat, as Tier 2 Fighters have a higher median Strength score of 18 and a median Dexterity score of 14. A new player torn between the two play-styles could take this data into consideration.



This entire project was primarily implemented through two main interactive data visualizer apps made with Shiny.

For the first app, 'DnD 5e Statistics Visualization', I initially based it on Shiny's 'reactivity' app, which I learnt about in Week 8 of this course, and modified it into an app that displays either a bar graph or histogram. The x-axis would be a DnD statistic (HP, AC, Str, Dex, Con, Int, Wis or Cha) chosen by the user, and the y-axis plotted frequency of various values for the chosen DnD statistic within a filtered dataset. This bar graph/histogram was plotted using ggplot2, a Week 7 concept. The user can choose whether they wish to print a bar graph or histogram with a radio button input, a new concept for me. If the histogram was selected, an additional setting that was previously not visible would reactively appear – a slider for setting the number of bins in the histogram. This was achieved using conditionalPanel(), another new concept. The bars in the graph itself could also be clicked on to reveal more information, another new concept I learned.

The filtering of the dataset was quite complex. The app has a panel with many customizations, allowing a user to choose what combination of DnD classes they wish to analyze, how to logically filter the dataset based on their class choices, the minimum and maximum level of characters in the filtered dataset and how sensitive outlier detection should be. Outlier detection was another new concept.

In order to allow for selection of classes, I had to tidy the way class data was stored in the original dataset. There are examples of players in the database who multiclass, resulting in their value under the 'justClass' variable displaying as 'Artificer|Fighter|Wizard' for example. To rectify this, I mutated the dataset to include new individual columns for each class in DnD 5e, the variable names being the class names. The values found under these columns would be either 'TRUE' or 'FALSE', indicating whether or not the player's character had that specified class. This was done using a for loop, a Week 6 concept, as well as grepl(), a new concept.

The logical filtering based on class choices had three options: 'Show characters with at least one of the selected classes', 'Show characters with only the selected classes' and 'Show characters with exactly the selected classes'. In the interest of word count, I cannot fully break down the method here, but it involves manipulating various logical vectors and calculating logical arguments row-wise using mutate(), transmute(), if_any() and if_all() and setdiff(). Mutate() was taught in Week 4 but the rest were new concepts.

The minimum and maximum level limits were slider inputs that filtered the dataset with two filter() logical conditions, a Week 1 concept. I tweaked it such that the minimum level can never be set higher than the maximum level and vice versa using observeEvent({}) and updateSliderInput(), two new concepts.

I also included a section that printed out the details of filtered data observations, high outliers and low outliers. The details were selected with select(), a Week 3 concept. The observations section had a 'page-flipping' system to allow users to easily view the entire filtered dataset.

The second app, 'DnD 5e Statistics Visualization - Exploring the Fighter', was mostly similar to the first but instead depicted the means, medians or modes (user's choice) of the six ability scores of Fighters in specific brackets of levels on a bar graph. The x-axis was the ability scores and the y-axis was the chosen summary statistic. Making a bar graph depict two variables instead of just one variable and its frequencies was a new concept. This app also used a custom mode() function, as R does not have a default function for calculating modes. Custom functions were taught in Week 5. The method for creating a mode function was a new concept.

References

Are these stats overpowered? Im new to DND and my DM told me to change them. Reddit. (2023, June 25). https://www.reddit.com/r/DungeonsAndDragons/comments/14if2oe/are_these_stats_overpowered_im_new_to_dnd_and_my/

Character progression planning, how do you do it?. EN World Tabletop RPG News & Reviews. (2019, November 11). https://www.enworld.org/threads/character-progression-planning-how-do-you-do-it.668394/

Oganm. (2022, August 24). *dnddata.* GitHub. https://github.com/oganm/dnddata

Wizards of the Coast. (2023a). *Fighter.* D&D Beyond. https://www.dndbeyond.com/classes/fighter

Wizards of the Coast. (2023b). *Player's Handbook: Dungeons & Dragons.* D&D Official | Dungeons & Dragons. https://dnd.wizards.com/products/rpg_playershandbook

Final submission word count: 1298