Week 9 - 11 Diary Entry

Yang Juan Hun

2023-11-10

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
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	

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.