Week 9 & 10 Diary Entry

Yang Juan Hun

2023-10-27

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.