Analyzing the Combos and Killpower for Smash World Tour Online Qualifiers

Report, Data Collection, & Statistical Analysis completed by bluesclues(Twitter: @jordanshluk, Reddit: /u/duckhunttoptier)

LINK TO SPREADSHEET DATABASE

Summary/Abstract

Smash World Tour's Online Qualifier provided a large sample of character combos and kill percents. I analyzed every recorded combo and stock taken, and used statistical analysis to determine ways we can classify, rank, and compare character's combo game and killpower. Results for combos indicated a correlation between average combo output and combos per stock ratio, thus leading to the potential to both score and classify combo games. Results for killpower were unable to cleanly organize characters into different archetypes, but different methods for ranking killpower were showcased. Factoring in kill confirms and hits per combo would help classify and rank character's and understand the metagame more.

Introduction

The Smash World Tour Online Qualifier was a series of online events for Super Smash Bros Ultimate(SSBU or Ultimate) taking place between February 1st through May 22nd, 2021. These online qualifier events determined which players would move forward to offline, regional stage two qualifiers. Selected Top 32 matches were streamed on twitch.tv/vgbootcamp for the public to spectate.

Using the streamed matches, I collected data for every combo performed, and every stock taken. Using the data collected, I sought after the following goals:

1. Could combo games be classified into different archetypes?
2. Which characters had the best combo game on wifi?
3. Could killpower be classified into different archetypes?
4. Which characters had the best killpower on wifi?
5. Can we find correlations between different numerical values to learn more about how Ultimate works as a competitive game?

This analysis will examine combo output and killpower independently, and will not focus on linking the two together.

Sampling Methodology

Using the recorded sets from VGBootCamp, I took each set at a time and recorded on a spreadsheet three different values: The opponent's weight, the damage each combo dealt on the opponent, and the percent at which the opponent lost their stock.

For characters who changed their weights (Pokemon Trainer, Pyra & Mythra, etc.) The average weight was chosen to be recorded. For Pokemon Trainer, that value was 96, and for Pyra & Mythra, the value was 95.

Determining combo output requires a lot of specifics, as the definition of “combo” can be either strictly, or loosely applied. For this survey, we aimed to keep the definition loose, and the following guidelines were put in place:

- Calculations were done by rounding up.
    - For example, if the starting percent of the opponent was 16.4%, and the combo finishes when the opponent was at 37.5%, the values are rounded to 16% and 38%, thus the combo did a total of 22%.
- Must be a minimum of two, individual moves that deal both damage and knockback.
    - For clarity, two different palutena neutral airs count as a combo. However, two hits from a singular neutral air does not count as a combo.
- If an opponent uses an escape option, but the advantageous player succeeds in following up on the escape option with another attack, the combo continues.
    - Escape options are defined as non-attacking inputs the player can perform. This includes things such as teching, air dodging, and special escapes such as Pokemon Swap.
        - For escape options that only take partial damage(Bat Within, Foresight, etc.), the combo will not count unless it is over two unique hits.
    - Catching jumps only continues to combo if the opponent hasn’t moved far from where they initially were sent.
- Trading with the opponent does not count as a combo.
    - If the first hit of a two hit combo was a trade, the combo is not counted.
    - If the second hit of a two hit combo was a trade, the combo is not counted.

        - If the first hit of a three(or more) hit combo was a trade, the combo is counted except for the first hit.
        - If the last hit of a three(or more) hit combo was a trade, the entire combo is counted.
    - POISON CLAUSE
        - Poison: Any attack that deals continuous damage post hitstun. Examples of these attacks include Jigglypuff's Rest and Joker's Eigan & Eiha.
        - If a poison attack is used to finish off a combo(implying the attack deals hitstun as well), then the poison damage will be added up until the opponent gets hit again.
        - If an opponent is taking poison damage while being combo'd by a combo independent of the poison, then the poison damage that occurs during the combo is counted up until the final hit of the combo.
    - PROJECTILE CLAUSE
        - If a two-hit combo that includes a projectile occurs, the combo will only count if:
            - The projectile was the *first* hit of the combo.
            - The projectile was clearly intended as a follow up.
            - Examples of a clearly intentional projectile followup include Robin's Arcfire to Thoron.
            - Example of a clearly unintentional projectile followup includes Young Link's dash attack into the returning hit of Boomerang.
        - All three-hit combos that include projectiles count, no matter what chronological hit the projectile was.
    - BLAST ZONE CLAUSE
        - If the opponent is sent into the blast zone and takes an additional 1.2%, the additional percent is added towards the total combo. The blast zone percent can only be added once per combo.

Here are the guidelines used for killpower:

- The percent before the opponent was dealt the final blow is what is recorded.
- If the opponent was killed with a kill confirm, the initial percent before the kill confirm(combo) began is what is recorded.
- If the opponent was killed with an edgeguard, the percent recorded is the hit that was dealt when the opponent was either:
    - Last on stage
    - Last above the stage

        - Last touched the ledge.
            - Two Framing counts as touching the ledge
    - SELF DESTRUCTION CLAUSE
        - If the opponent self-destructs, then the value is not counted.
        - Self destructs are counted as clear indications of opponent's taking their stock without the other player's actions and inputs playing a role in their loss of stock.
            - Missed recoveries are counted by subjectivity. If the opponent was clearly capable of recovering, but failed to due to a buffered airdodge or a poor angle, then it is considered a self destruct. However, if the recovery was clearly difficult and required precise timing and spacing, then the stock can be counted if they fail to recover.

After each set was recorded, the following values were analyzed for each character's killpower and combo output.

- Average
- Median
- Lower & Upper Quartiles
- Skewness
- Standard Deviations

Additionally for combo output, a combo per stock value(CPS) was counted, dividing the number of combos a character landed by the number of stocks the character landed. A CPS for combos with high damage output(above 40%) was also recorded. A value table would determine how frequent a range of combo output's was. The same table, though different value ranges, was done for killpower as well.

Lastly, the total # of combos, and total # of games for each character was listed on. All character data was kept on an individualized spreadsheet. Figure 2 shows an example of a character's spreadsheet. For comparisons between characters(final scoring), only those with ten or more games recorded were used.

OPP. WEIGHT	COMBO	KILL								
91	42	96								
91	10	168	Average Combo	28.11111	IQR	13	Average Kill	122.0512	IQR	59.5
91	23	124	Median	25	Q3	33	Median	124	Q3	153
91	70	91	St. Dev	13.4939(	Q1	20	St. Dev	48.9075	Q1	93.5
	46		Average Weight	101.4838	# of Games	19				
	11		Combo Spread	%	CPS	3.692				
	30		0-10%	2.78	HCOR	0.181				
	15		11-20%	25.00	Total Combos	144				
	28		21-30%	42.36	Best Combo	77				
	13		31-40%	11.81						
	13		41-50%	11.11						
	10		51-60%	2.78						
	22		61%+	4.17						
	26		Kill Spread	%						
	10		0-25%	7.69						
	15		26-50%	0.00						
	34		51-75%	7.69						
	13		76-100%	17.95						
	26		101-125%	17.95						
	20		126-150%	23.08						
	49		151-175%	12.82						

Figure 1. Example template of character data collection.

Analysis Procedure & Results

General Results

In total, almost 10,000 combos(9,830 exact) were recorded, and 2,028 games were analyzed. The process took approximately 90 hours total.

53 characters had enough data to analyze and compare. Of those 53 characters, only one character had over 100 different games analyzed(R.O.B). Figure 3 showcases the ranking of character sample sizes(# of games streamed).

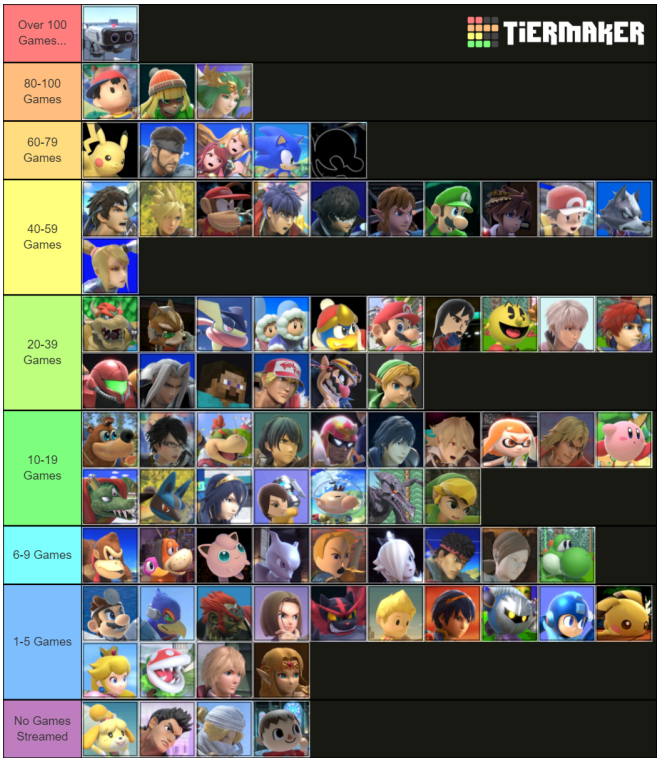

Figure 2: Character sample sizes. Those above 10 games were considered for comparisons and rankings. Not ordered past the first three tiers. N = 2028

Combo Results

First type of analysis to score and compare combo games exclusively looked at average damage output. Figure 4a ranks each character using this methodology. Bowser had the highest average, at 34.28%. Pikachu had the lowest average, with 21.59%. The median value was 25.83%, with Roy being the middle of the sampled characters.

To determine if the opponent's average weight had an impact on the sampled character's average, we tested a correlation, as showcased in figure 4b. There is likely no correlation between average weight and average combo output.

Figure 3a: Average Combo Output. N = 53. Maximum = 34.28%(Bowser). Minimum = 21.59(Pikachu). Median = 25.83(Roy).

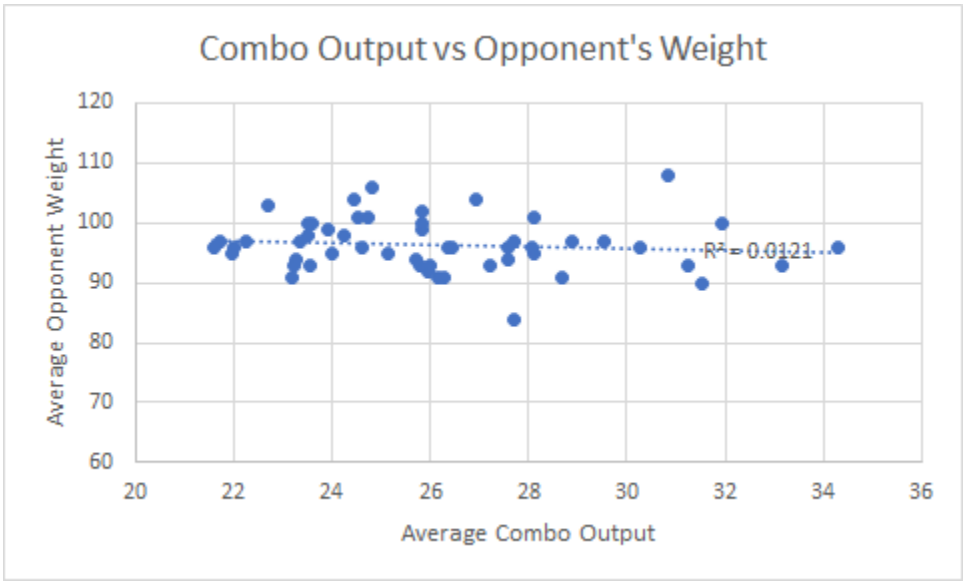

Figure 3b: Combo Output vs Opponent's Weight. R^2 = 0.0121. There is likely no correlation between average combo output and average opponent weight. N = 53

After examining just the average combo output, I determined there were likely other factors we could use to classify combo archetypes. I looked at the correlations between average output and combos per stock (Figure 4c), and also correlation between average output and interquartile range (Figure 4d). Additionally, I also looked to see if there was a correlation between CPS and Range (Figure 4e).

The results indicated that there were clear relationships with combos per stock and combo output, a potential relationship between combo output and range.

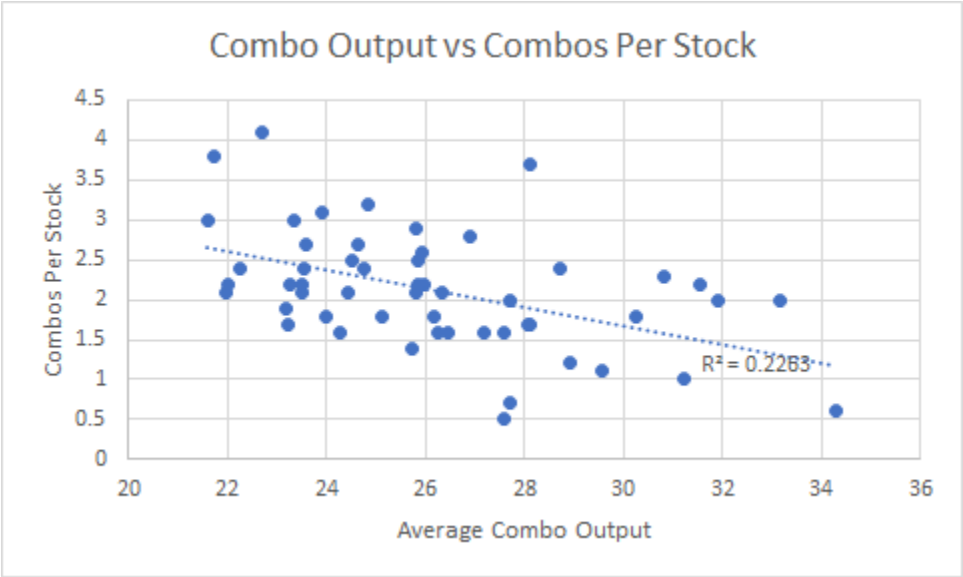

Figure 3c: Combo Output vs Combo Per Stock. R^2 = 0.2263. Evidence indicates a correlation between Average Combo Output and Combos Per Stock, where higher damage output indicates a lower CPS ratio. N = 53

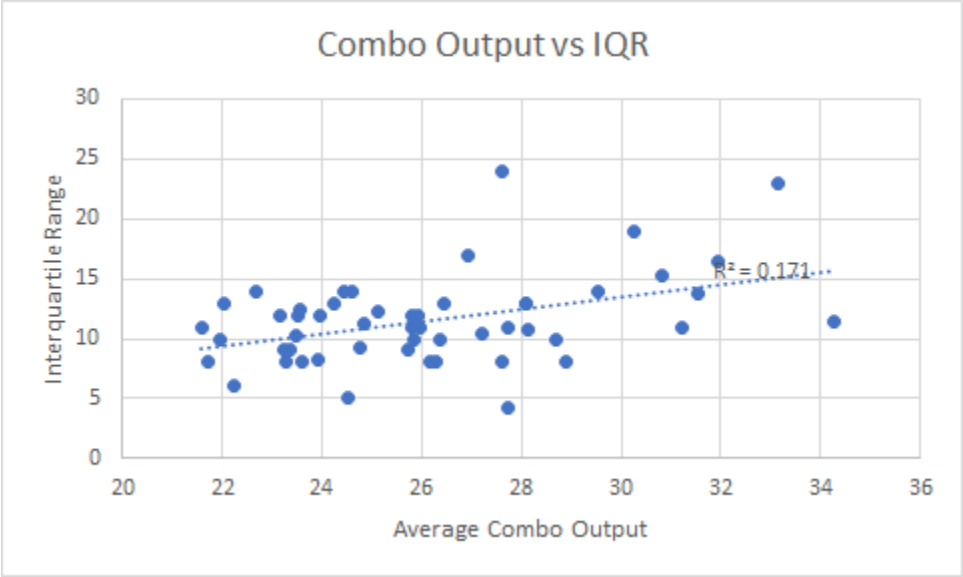

Figure 3d: Combo Output vs Interquartile Range. R^2 = 0.171. Evidence indicates a loose correlation between Average Combo Output and Interquartile Range, where higher damage output indicates a larger interquartile range. N = 53

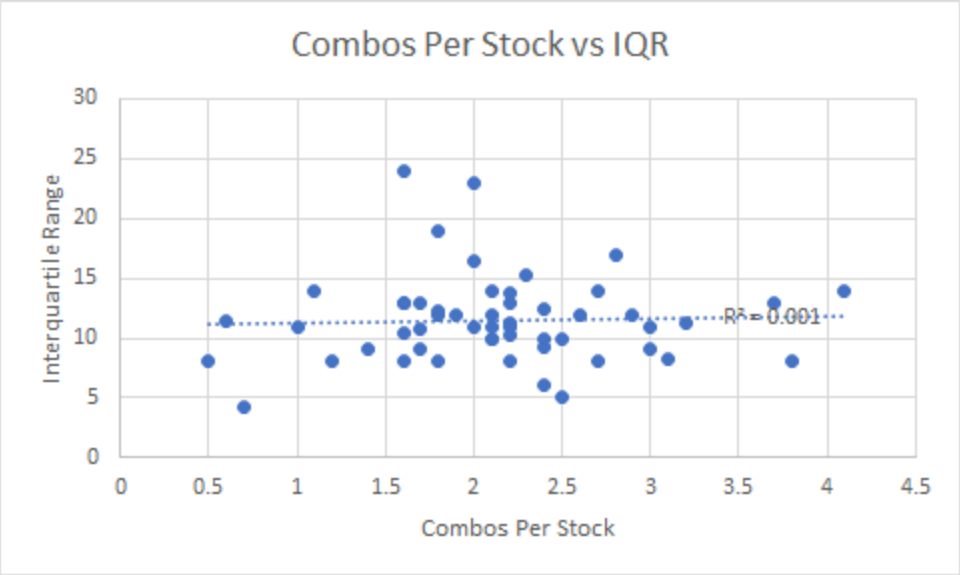

Figure 3e: Combos Per Stock vs Interquartile Range. R^2 = 0.001. Evidence indicates there is no correlation between combos per stock ratio and interquartile range. N = 53

Moving forward, I decided that all three values would be used to classify, but I would also use the relationship between damage output and combos per stock to create a score to determine the best combo games.

Figure 4f showcases the combo per stock ratios, ranked. The relative standard deviation for CPS was far greater than the %RSD value for average combo output(11.62 for damage output, 35.5 for CPS).

Figure 4g showcases the interquartile ranges of each character's data. As the variability of the combo output is subjective, as having a large range can have positives and negatives, interquartile ranges were only used for classifying archetypes.

Figure 3f: Combos Per Stock ratio for each sampled character. The highest ratio is 4.1(Ken). The lowest ratio is 0.5(King DeDeDe). The median ratio is 2.1(Chrom, Joker, Palutena, Link & Pokemon Trainer all have the median ratio). N = 53

Figure 3g: Interquartile ranges of sampled characters. The maximum range is 24(Olimar), while the minimum range is 4.25(Mii Gunner). The median IQR is 11. N = 53.

After collecting and organizing the average damage output, CPS, and IQR, characters were classified based on the correlation between CPS and damage output, using the archetype grid in figure 4a.

Assuming characters with higher CPS had lower damage outputs, characters who followed the correlationships were either heavy hitters(high average, low CPS), or abundance(low average, high cps). Those who broke the correlation by having high average and high CPS were considered rulebreakers, and those who had low

averages and low cps were considered either scrappy, or lacking. Those who did not fit a clear archetype due to being in the middle were considered intermediates.

Additionally, IQR was used as a secondary classification, to determine if the characters had uniform combos(IQR less than 12), or those who had a diverse array of unique combos(IQR greater or equal to 12).

Figure 4b is the overall result of combo classification.

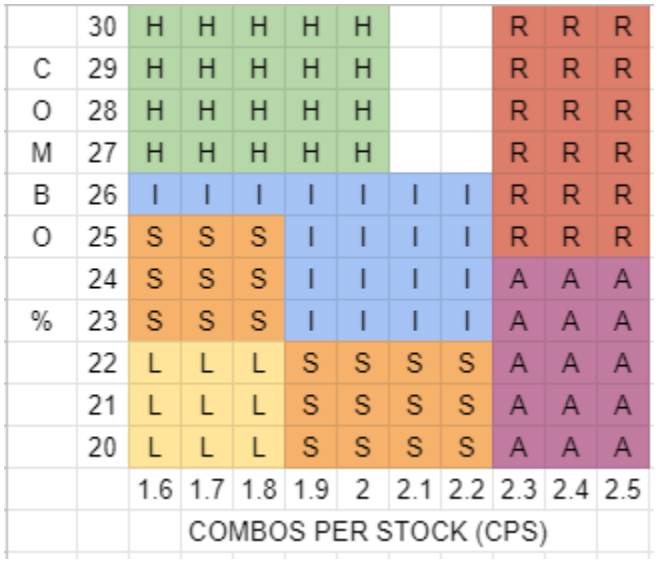

Figure 4a. Grid determining the different combo archetypes. R = Rulebreakers, A = Abundance, H = Heavy Hitters, I = Intermediate, S = Scrappy, L = Lacking. White space indicates no characters were within the area.

Figure 4b: Result of Archetype Classification for Combos. No order within tiers. N = 53.

To determine which characters had the best combo games a, each character was given a relative score based on the maximum value for each value. Those two values were added up, then weighted. I decided to weight average damage output more than CPS due to the high %RSD for a weighted result.

The weighted combo score formula:

((Character's Average/Maximum Average)*.625)+((Character's CPS/Maximum CPS)*.375)

The unweighted combo score formula:

((Character's Average/Maximum Average)+(Character's CPS/Maximum CPS))/2

Figure 4c is the weighted combo score, while figure 4d is the unweighted score.

Figure 4c: Weighted combo score. Highest Score = 0.85(Bayonetta). Lowest Score = 0.55(King DeDeDe). Median Score = 0.67(Ness). N = 53.

Figure 4d: Unweighted Combo Scores. Highest Score = 0.86(Bayonetta). Lowest Score = .46(King DeDeDe). Median Score = 0.63(Palutena)

Killpower Results

I decided to look at averages first, as the most clear indicator of looking at killpower would be the lowest average kill percent. I also chose to look at skewness, to determine if the characters kill earlier or later than their average. Figure 5a displays

average kill percent, while figure 5b shows the skewness of each character's kill power.

Figure 5a: Average Kill Percent. Best Average = 99.46(Lucario). Worst Average = 135.23(Richter). Median = 119.88(Mii Gunner) N = 53

Figure 5b: Skewness of Average Kill Percent. Negative Skew(< - 3.00) Indicates character killing earlier than their average. Neutral Skew(Between -3.00 and 3.00) indicates generally killing around the character's average. Positive Skew (>3 indicates generally killing later than their average. Unordered within tiers. N = 53.

After looking at averages, I decided to only test two relationships, one with interquartile range, and the other with weight. Figure 5c showcases that there could be a rough, positive correlation between averages and interquartile range, however nothing concrete, either.

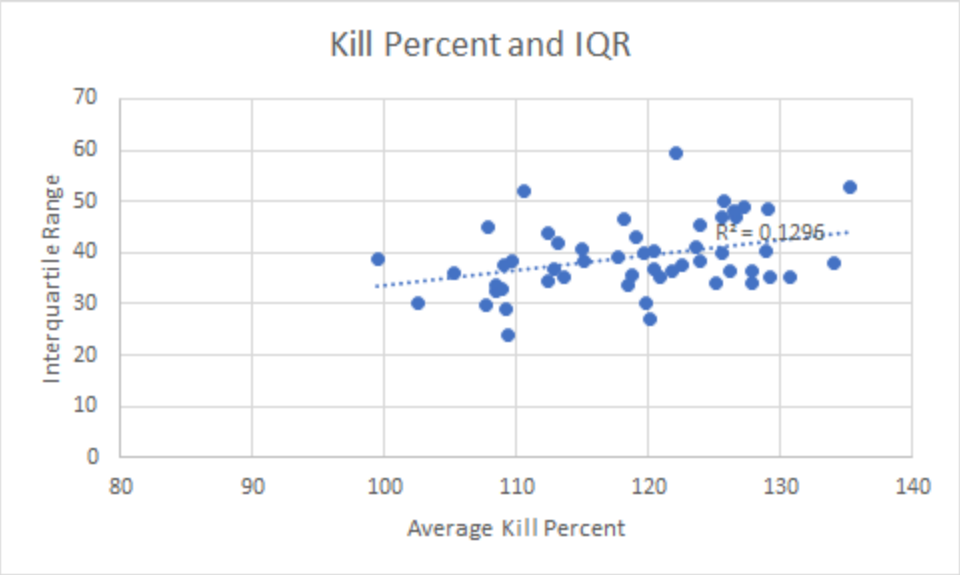

Figure 5c: Kill Percent and IQR. R^2 = 0.1296. Evidence may indicate a slight correlation between low average kill percents and small interquartile ranges. N = 53.

For determining the relationship between weight and kill percent, I wanted to first look at it through a specific character basis, then through a general basis. Figure 5d showcases average kill percent for each character and the average opponent's weight.

Secondarily, I wanted to test if there would be a relationship for certain characters between weight and kill percent. Thus, using randomly generated numbers, I selected 4 different characters to look at: Mario, King DeDeDe, Pikachu, and Wario. Figure 5e is a compilation of all 4 character's individual analysis of killpower and weight correlation.

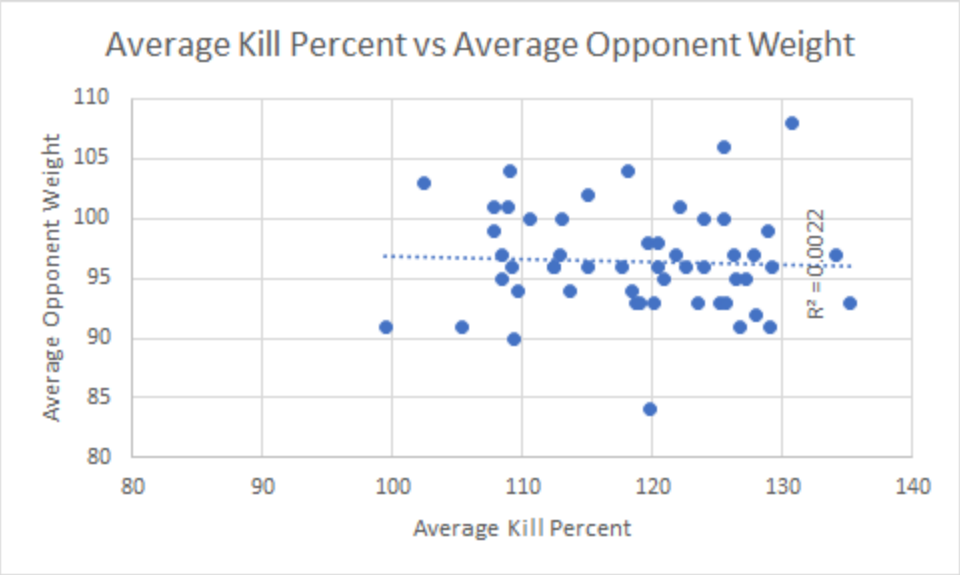

Figure 5d: Average Kill Percent vs Average Opponent Weight. R^2 = 0.0022. Evidence indicates no correlation between average kill percent and average opponent weight.

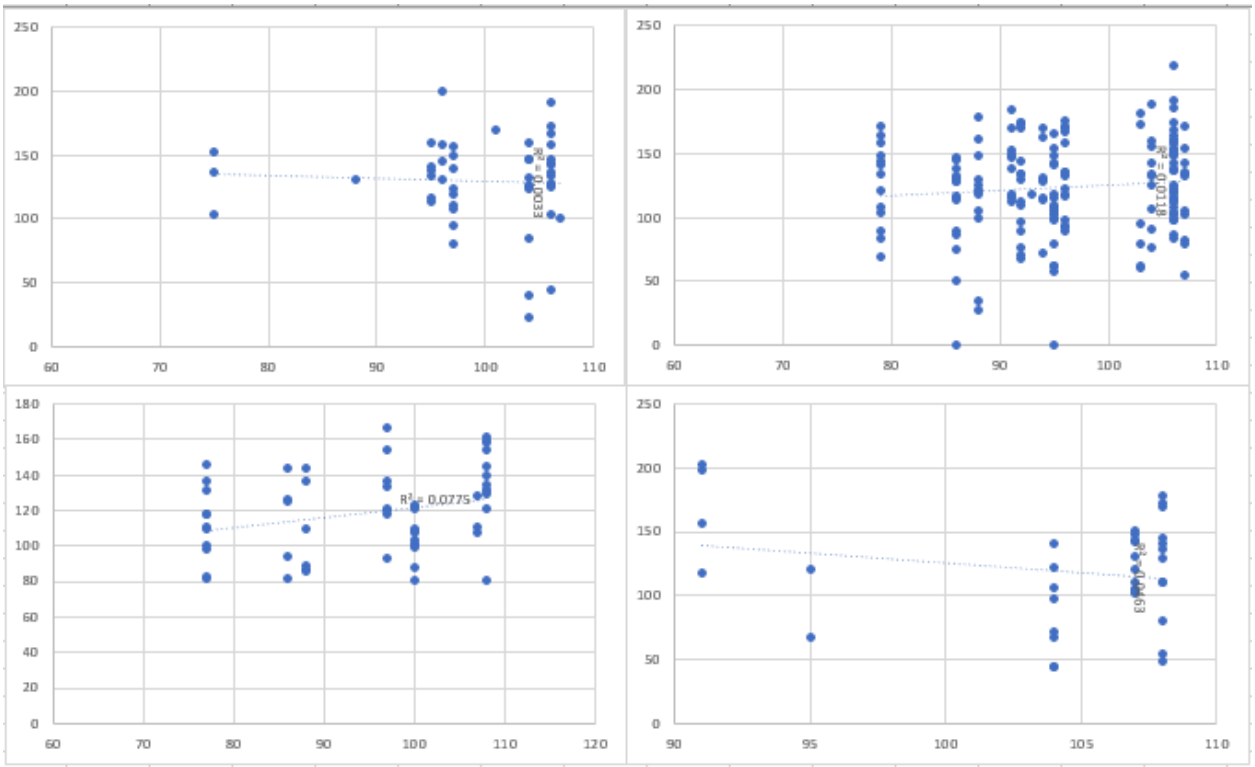

Figure 5f: Individualized relationships between Kill Percents and Average Weight Trendlines indicate that characters could have unique and different correlations and trends. N = 4

The last two values I examined for killpower were interquartile range, and a character's percent of struggle(in other terms, the percent of kills a character has past 150%). Figure 5g ranks each character's interquartile range, while figure 5h showcases the percent of time each character took stocks past 150%.

Figure 5g: Interquartile Ranges of each character's killpower. Largest Range = 59.50(Bayonetta). Smallest Range = 24.00(Inkling). Median Range = 38.00(Toon Link). N = 53.

Figure 5h: Struggle for Stocks past 150%. Lowest Value - 0%(Ken). Highest Value - 35.90%(Richter). Median Value = 15.05%(Link). N = 53.

Discussion & Conclusion

I believe there was some degree for each objective, except for classification of killpower archetypes.

Combo games could be classified pretty clearly into six different types, five of which present in the database. Factoring in combo diversity, this can be doubled into 12 different archetypes. A weighted and unweighted ranking displaying potential best combo games on Smash Ultimate Wifi were also presented, indicating that statistically ranking combo games can be possible, even offline.

While we were able to find different ways of ranking killpower, I was unable to find a way of classifying different archetypes, due to a lack of any clear numerical relationships. The last goal found success, as relationships were found between individualized characters' killpower and opponent weight, alongside combo per stock ratio's correlation with average combo output.

While there were successes, I believe there were multiple issues with the procedure and sampling methodology. Self destructs, and situations where combos occurred but the stock wasn't taken happened a significant amount of time during sampling, and because the CPS ratio wasn't a separate value and rather calculated through formulas, it isn't entirely accurate.

Additionally, in future methods, I would recommend looking at the percentage of combos that took stocks(kill confirms), and also the number of hits per combo. Luigi and Cloud are both considered Diverse Heavy Hitters, however it is clear that luigi's combos have multiple actions while Cloud's generally consists of only two hits per combo.

When it comes to killpower, I would like to see future work done on how characters get kills as opposed to what percent they were at. Seeing what percent of a character's kills are edgeguards, ledetraps, whiff punishes, out of shield options, and so forth would likely play a more interesting role in determining killpower archetypes than just numerical values.

I highly recommend that this type of data be applied to players individually, rather than just a large group of characters. I would be highly interested in seeing the sampling and analysis applied to comparing different players who use the same character, and will likely apply this method for this.

I look forward to seeing how the project is applied to future analysis and studying of Smash Ultimates metagame. Thanks for reading!