

Data Entry

The data are taken by hand like so:

32	F4b	Jon Snow (z)	repair manner male?	g) Masa ✓ 12:46 (149)	King C
I	II	hatch gas pool hatch		h	gas rax cc fact
I	I	z dro Q		h	reap heli
III	I	Q spo ors dro bone z hyd		h	5 av
II		hyd z Q bone dro vip ors		p	bsh holi tank hbat
32	F5a	Jon Snow (z)	workslighter	Astrea g) V 17:23 (150)	King C
III	III	hatch gas pool hatch			nexus gate gas gas
II	I	z		h	z ad st all
II	I	Q spo dr hyd mut		h	or
II	I	bone z hyd vip Q ors corr		p	co ll arch st dos z se HT
II	II	hyd bone z corr vip ors bl		h	call arch st dos z se corr AT
I		hyd z dro		h	z wp
32	F5b	Jon Snow (z)		2:54 ✓ Astrea (p) (151)	portalex
I		hatch gas pool hatch			gate gas nexus cyber
II		Q z		h	ad
I	I	h z		h	st ad
I	II	Q z b		h	wp z
I	II	ro ors Q		p	wp z obs
I		Q dro		h	wp HT
I		ro ors		h	imm z se arch
I		ro rax Q dro		p	imm z se arch up

- First Row Elements

- 1) Round
- 2) Group
- 3) Match number
- 4) Game
- 5) Name of First Player
- 6) Race of First Player
- 7) Game Tags
- 8) Race of Second Player
- 9) Name of Second Player
- 10) Winner Checkmark
- 11) Length of Game
- 12) Game ID (indicates data have been entered onto the computer)

13) Map

Second Row (Opener) Elements

- 1) First four non-supply buildings made by Player 1
- 2) First four non-supply buildings made by Player 2

Third and Later (Encounter) Row Elements

- 1) Attack Count
- 2) Rally Count
- 3) Player 1 Interaction Type
- 4) Player 1 Units
- 5) Player 2 Interaction Type
- 6) Player 2 Units

Method of initial data entry into a spreadsheet:

ROWID	GameID	Round	Group	Match	Tag	P1	P2	P1Race	P2Race	Winner	Map	Minutes	Seconds	round	group	match	game
9952	214	0			information	time	silky	t	z		2 kairos jun	29	3	16 d			1 b
9953	214	0			game tags	a slow death											
9954	214	1			count			opening									
9955	214	1			time	barracks											
9956	214	1			time	refinery											
9957	214	1			time	command center											
9958	214	1			time	factory											
9959	214	1			silky	pool											
9960	214	1			silky	hatchery											
9961	214	1			silky	extractor											
9962	214	1			silky	hatchery											
9963	214	2			count	1	2 harassme	time									
9964	214	2			time	reaper											
9965	214	2			silky	zergling											
9966	214	2			silky	queen											
9967	214	3			count	1	0 harassme	silky									
9968	214	3			time	scv											
9969	214	3			time	marine											
9970	214	3			time	reaper											
9971	214	3			silky	zergling											
9972	214	3			v	1	0 harassme	time									
9973	214	3			time	medivac											
9974	214	3			time	hellion											
9975	214	3			silky	queen											
9976	214	3			silky	zergling											
9977	214	3			silky	drone											
9978	214	4			count	5	1 harassme	time									
9979	214	4			time	hbat											

After data entry I clean the data up carefully and split it into game/engagement/unit/tag tables, so that it is organized somewhat like a relational database for easy use by Mina, who, using Tableau, creates the visualizations for our articles.

A More Thorough Explanation

I designed this data entry to fill two requirements:

- No replay is needed since not all tournaments release them;
- Taking data could be done in real time as I watched games.

Why take data by hand?

- I only have one monitor and no tablet
- I always have taken data by hand for scientific research, so it feels comforting and ‘right’ to me.

Game Information



This is simply the overall game information, all of which aside from the tags and time could be taken from Liquipedia if I wanted to.

Game Tags: It takes little effort to make notes on the games I think are interesting, and maybe somebody would want to know which matches had, say, a legitimate base trade. I make up tags on the fly about unique events (muta balls trade perfectly evenly), strategies (mass cyclone tech swap), special intensity (“slugfest”) or player interactions (banter). Less than half of the games have any special tags.

Game Duration. Human reflexes are slow, especially when you are still learning a method. I found I could not rely on recording the moment of typing “gg” or the concede. I mark the time of the next second change after the ‘game won’ animation completes, and subtract 2 seconds from that time during the final cleanup of the computerized data.

Opening Buildings



The first line after the game information has the first four non-supply buildings made. This is supposed to discriminate between different openers in a way that's easy to quantify. I can't always count on commentators or observers to tell me exactly what opener people use, and I am afraid to rely on a simple description of them because openers could theoretically change over

time. After about four buildings get built, the game tends to get far more chaotic and keeping track of build orders becomes difficult for a layman like me; and finally, it would be difficult to track the exact timing of buildings for both players without a lot of practice. Therefore, I only record the sequence of first four non-supply buildings made.

Engagements

32	F4b	Jon Snow (z)	repair minion male?	or Masa ✓ 12:46 (149)	King C
I	II	hatch gas pool hatch		gas rax cc fact	
I	II	z dro Q		h reop heli	
III	I	A spo ors dro bone z hyd		sqv bsh	
II		hyd z dro bone dro up ors		bsh holi tank hbat	

Each additional line represents a series of similar engagements.

Engagements counted together on a line are initiated by the same player and have related, but not necessarily exactly the same, unit compositions. Early on in planning this data-taking method I recognized that writing down an entire unit comp of up to several unit types for each engagement that added just one new unit type would either reduce my attention to the game, causing me to miss something critical, or force me to pause it regularly, so that it would be even more time-consuming to take data. I decided to count similar engagements, and when a new unit is added to the composition, I just add it to the same line with an underscore (entering the underscore as a note in a LateAddition column of the spreadsheet). NOTE: players often remove units from their army compositions as well, but I do not notate that. I would not recommend assuming that each listed unit is involved in all engagements of its type even if it isn't a late addition.

Usually, the aggressor in the engagement is obvious. Sometimes both players move out at once and meet in the middle. If one army has moved farther, I call them the aggressor; otherwise I just pick one of the two at random/based on my gut. This situation is rare and should not affect the data much in any case.

Not counted as engagements:

- *Killing of scouts*, unless they take potshots at the enemy during scouting.
- *Killing of creep tumors*. I was not convinced I could keep track of whether or not creep tumors were killed in large engagements, so I ignore them completely.

- *Intentionally sacrificed workers*. Intentionally sacrificed army does count, though, assuming it attacks; and sad mistaken move-commands do as well since they are costly.
- *Casts of revelation*, since this is scouting, not aggression.

Counts

The first two columns of data entry show tallies of the number of engagements of that type.

- *Attacks*, the first column, shows how many times the aggressor moved from their side of the map to confront the enemy.
- *Rallies*, the second column, shows how many times the aggressor attacked after a move-out listed in the first column, but instead of going home afterwards, they sat on the opponent's side of the map to regroup or accept reinforcements before attacking again.

I split these because it tells you a bit about the army movements in an easily quantifiable way. Rallies can occur after as short a time as stepping back out of range for a few seconds and then moving back in, or as long a time as minutes of a soft contain. Attempts to break a contain are generally counted as aggression by the player holding the contain, though, since they started it, even though this doesn't make intuitive sense.

Types of engagements

The initiator of an engagement is marked in a small column with a H, P, or X indicating one of the following actions:

- *Harassment* is aggression using a small number of units or only part of the army. There was a short period of time where I counted early-game attacks with more than 10 supply as not harassment, but I decided this didn't entirely make sense. I now rely on my gut, which depends on years of watching Starcraft 2, to decide whether a specific engagement is harassment or not. Hopefully that is good enough for you.
- *Pushes* are your standard move-out using a large part or entirety of the army.
- *Proxies* are buildings made outside the normal bounds of a base. Attack/rally counts of proxies indicate how many of the building was made and whether (as in the case of pylons or nydus worms) they were made for the same attack (in which case additional ones are counted in rallies) or different attacks (all counted in attacks). Defending units against proxies are noted only if those units arrive and start damaging the proxy before it serves its purpose (makes a unit, or finishes for nydus/tech buildings).

NOTE: Similar types of harassment or pushes taking place at the same time in two different places are counted as one. For example - if someone flies two drops from the opposite sides of their opponent's base at once, this is counted as one harassment instance. Why? It's a little easier for me to keep track of.

Units

The units used by each player (both aggressor and defending) are listed, abbreviated for speed of writing and space. Underlined (“LateAddition”) units were added to the army composition after the first attack, but there is no notation for units removed from the composition. The list includes detector units and battle relevant buildings like shield batteries, bunkers, and spore crawlers.

Not listed:

- Creep tumors
- Changelings
- Non-attacking/healing buildings, even if killed
- Hallucinations
- Overlords, unless the pointed object of harassment or serving as dropolords
- Mind-controlled units that have changed sides

Are Listed:

- Nukes, because everyone loves them, they will have their own line and counts because this is not hard to do