MLG COLUMBUS 2016: AN ECONOMETRIC ANALYSIS OF COUNTER-STRIKE

By Christopher Sardegna

INTRODUCTION

eSports — a form of sports where the primary aspects are facilitated by electronic systems — provide unique opportunities to mine and analyze data. Every aspect of eSports matches can be recorded digitally and analyzed with software, vastly reducing the energy required to collect and interpret patterns. The recent MLG Major Counter-Strike Global Offensive competition in Columbus, Ohio is one of these opportunities.

<u>Counter-Strike: Global Offensive</u> (CS:GO) is an internationally acclaimed competitive skill-based First Person Shooter developed by Valve LLC. Released in 2000, it has a long history in competitive gaming (eSports), attracting over a million viewers at the 2016 Major League Gaming competition in Columbus, Ohio.

Game Overview

In competitive mode, matches are broken down as follows:

- Two 15-round halves (30 possible rounds without overtime)
- 16 round wins required to win (more if overtime has been reached)
- The first round of each half (and each overtime) restricts weapon purchases to pistols
- Depending on the outcome of the round, each team is rewarded with in-game money to purchase better gear
- Winning the pistol round guarantees a minimum \$3000 reward while losing nets \$1400

Research Question(s)

Originally, this project's aim was to ascertain the impact winning a pistol round in Counter-Strike: Global Offensive has on winning the match. As the data grew, the scope of this project increased beyond this single question to a general analysis of the effectiveness of teams' economic management.

DATA PROCESSING

<u>HLTV.org</u> keeps a brilliant searchable <u>index</u> of historical CS:GO matches. To focus on the Columbus 2016 competition, set *Events* to *MLG Columbus 2016* and download the demo files for each match. Demo files contain highly compressed information about matches. They allow users to accurately recreate matches inside the game engine. Demo files also provide a unique window for analyzing play without manually recording a single piece of data.

Parsing the Demo

<u>CSGO-Demos-Manager</u> is an invaluable open-source tool for analyzing CS:GO demo files. It will "play" through a demo, tracking many different variables. Because the project is open source, it is easily extensible. However, it only runs on Windows.

CSGO E	DEMOS MANAGER						Setting	5 _		×
17 An	alyze	₽ Refresh 🖉 Browse 💼 Delete 😂 Stats	Export	Account stats:	👻 🖌 Show all					
► Wate	h 🧲 Goto Tick	🕹 Watch highlights 🏾 🎔 Watch lowlights 🛛 🖑 Add to suspects	😹 Suspects (0)						
Folder:	c:\program files (x86)\st	eam\steamapps\coi 🔻 🔄 Show all folders 🖌 POV 🖌 Valve 🗸 eBo	t 🖌 Faceit 🗸	esea 🖌 cevo 🖌 p	opFlash 🖌 No analyzable	e demos 🛛	ilter		;	×
Source	Date ↓	Name	Мар	Team 1	Team 2	Status	ST1	ST2	Ban	
VALVE	2016/04/10 19:13:00 N	/LGColumbus2016-natusvincere-vs-luminosity-bo3-map2-overpass.der	n de_overpass	Luminosity Gaming	Natus Vincere					ĺc
VALVE	2016/04/10 19:13:00 N	/LGColumbus2016-natusvincere-vs-luminosity-bo3-map1-mirage.dem	de_mirage	Natus Vincere	Luminosity Gaming					c
VALVE	2016/04/10 19:12:45 N	/LGColumbus2016-natusvincere-vs-astralis-bo3-map2-dust2.dem	de_dust2	Astralis	Natus Vincere					c
VALVE	2016/04/10 19:12:45 N	/LGColumbus2016-natusvincere-vs-astralis-bo3-map1-inferno.dem	de_inferno	Natus Vincere	Astralis					c
VALVE	2016/04/10 19:12:45 N	/LGColumbus2016-natusvincere-vs-nip-bo3-map2-mirage.dem	de_mirage	Ninjas in Pyjamas	Natus Vincere					c
VALVE	2016/04/10 19:12:44 N	/LGColumbus2016-natusvincere-vs-nip-bo3-map1-inferno.dem	de_inferno	Natus Vincere	Ninjas in Pyjamas					c
VALVE	2016/04/10 19:12:20 N	/ILGColumbus2016-liquid-vs-luminosity-bo3-map2-cache.dem	de_cache	Team Liquid	Luminosity Gaming					c
VALVE	2016/04/10 19:12:20 N	/ILGColumbus2016-liquid-vs-luminosity-bo3-map1-mirage.dem	de_mirage	Team Liquid	Luminosity Gaming					
VALVE	2016/04/10 19:12:19 N	/LGColumbus2016-mousesports-vs-nip-bo3-map3-overpass.dem	de_overpass	s Ninjas in Pyjamas	mousesports					
VALVE	2016/04/10 19:12:19 N	/LGColumbus2016-mousesports-vs-nip-bo3-map2-cache.dem	de_cache	mousesports	Ninjas in Pyjamas					
VALVE	2016/04/10 19:12:19 N	/LGColumbus2016-mousesports-vs-nip-bo3-map1-cobblestone.dem	de_cbble	mousesports	Ninjas in Pyjamas					
VALVE	2016/04/10 19:12:19 N	/LGColumbus2016-g2-vs-virtuspro-bo3-map3-cobblestone.dem	de_cbble	G2 Esports	Virtus.Pro					
VALVE	2016/04/10 19:12:19 N	/LGColumbus2016-g2-vs-virtuspro-bo3-map2-inferno.dem	de_inferno	Virtus.Pro	G2 Esports					C
VALVE	2016/04/10 19:12:19 N	/LGColumbus2016-g2-vs-virtuspro-bo3-map1-train.dem	de_train	G2 Esports	Virtus.Pro					C
VALVE	2016/04/10 19:12:19 N	/LGColumbus2016-luminosity-vs-virtuspro-bo3-map3-overpass.dem	de_overpass	s Virtus.Pro	Luminosity Gaming					C
VALVE	2016/04/10 19:12:19 N	/LGColumbus2016-luminosity-vs-virtuspro-bo3-map2-cobblestone.der	n de_cbble	Luminosity Gaming	Virtus.Pro					C
		Show more demos						42 d	emo	(s)
Comme	nt :				To watch	n 🔲 Wato	hed 🗖	Not a	halyza	ble

CS:GO-Demos-Manager

Once CSGO-Demos-Manager has analyzed the relevant demo files, users can choose to export the data to Excel as a group or individually. Both of the exports look more or less the same: a 6-sheet Excel table:

Δ	8	c	DEE	6 H I		к		м	N	0 9	0	8	5	T	ш	v	w	X Y	2	64	48	K I	10 A	ur al	- 10		ни		л×	AL 4	м	AN I	10	P 10	48	45	AT
Demi	NumTin	*	Duratii Winnei Winn	elEnd reas Type Side	Team Ki	ills 1	K 2	K 38	4	58	Trac	te Llum:	LADP	TDH 1	TDA F	lomb I Br	mb i Bo	omb Start r	Faulaer	Start m	FouiperF	lashb Sr	toke HE	Dec	ov Mole	to ince	and Match Winle	Team 1Doe	es Tes D	oes Tea Sta	t Tea S	Spend Te Ro	und CT	or T Win	Differe	en Winne	Differe
de.04	1	2700	164 Lumino CT	Terrorists Pistol round	10411 10	7	4	0	1	0	0	1	99.4	960	44	0	0	0 4000	4350	4000	4250	2	2	0	0	0	0 Luminosity	1	0	1	0	100	1	1	o	0 1	100
de or		23558	90 Lumino CT	Terrorists Semi-E-T	Natus								84.3	737	106			0 18400	18600	8150	7800						0 Luminosity			1 1	0260	10800			1 1025	0 1	10800
de_or		35182	65.3 Lumino CT	Terrorists Eco T	Natusiv								2 79.4					0 20750	21600	9850							1 Luminosity				900				1 1090		20200
de_o+		43548	97.8 Lumino CT	Counter-Tillionnal									112	979				1 34300	32150	21450	21350						2 Luminosity			1 1	2860	10600			1 1285	0 1	10800
de_or		56070	94.9 Lumino CT	Terrorists Normal									127	1136	198			0 36650	30550	20600	20750	6					1 Luminosity			1 1	6160	5800			1 1615	0 1	9800
08,01	8	63225	75.1 Lumino CT	Terrorists Force b T	Natusk								126	1156	107			0 38450	\$2050	18950	18600					0	0 Luminosity			1 1	3500	13450			1 1960	0 1	13450
08_01		77661	92.7 Lumino CT	Terrorists boo T	Natusia	2	3			2			68.9	805	83			0 30850	39060	16750	8900	8					3 Luminosity				2100	24150			1 1210	2	24150
200	0	107114	293 Natura VI	Unknown Normal		- 2	- 2							1992	20			0 34900	33250	20150	20400	12		-	0		3 commonly				2000	0000		2	1 -1100		-0500
200	-	120109	161 Natura VI	Creation 7 Normal		÷.	3						110	1048	107			0 32500	30750	20000	20300		2				 Commonly 				2000	4400			0		10100
1.0	11	160505	119 Lumina CT	Terrorists Normal						õ	0		80.0	205	91	ě		0 20600	26750	22900	23350	10	ě				1 Luminosity		ő		2300	3400			0 .230	ő i	3400
de or	12	154943	90 Lumino CT	Terrorista Doo T	Nature V			0			0	ò	83	762	60			0 18650	26550	10350	1050			4		0	1 Luminasity				300	255550			1 630	ó i	25550
de av	13	116464	135 Lumino CT	Terrorists Normal									126	1123	133			0 29400	20700	21050	21500		9				3 tuminosity				0250	8900			1 685	0 1	8900
de_or		213792	98.1 Lumino CT	Terrorists Doo T	Natus V								54.7		66			0 38350	31750	15950	10200						5 Luminosity			1 2	8400	20150			1 2240	0 1	20650
de_av		226151	202 Lumino CT	Terrorists Normal									22.6	687				0 31000	31700	21850	22550						3 Luminosity				9150	9150			1 915	0 1	9150
de_av	16	252160	141 Lumino T	Unknown Platol round									D 147	1242	224			0 4000	4350	4000	4400						0 Luminosity					-50				0 0	-50
de_or	17	270145	138 Lumino T	Unknown Force b T	Lumino		5				0		1 112	1005	117			0 8800	18950	18700	8650				0	0	0 Luminosity		0	1 -	3300	10100		0	0 -990	0 1	10100
de_01	16	287847	62.5 Lumino T	Unknown Force b T	Lumino							4	2 124	1104	133			0 10850	19650	19100	10900						0 Luminosity				6250	8950			0 -825	9 1	6950
00.m		2096	138 NEUSVOT	Counter-TP listol round							0		2 139	1408	161			1 4000	4300	4000	4300					0	0 cuminasity					100760			1 440		00
		20168	122 Matura VCT	Terrorists biormal	Contraction	- 1				š	0		101	1010	159	š	š	0 19750	26260	21450	22200				0		3 Luminosity				2200	3650			1 -220	ő ő	3650
de m	4	45903	86.3 Natus VCT	Counter-TEco T	Lumino	ě.	ŝ	ő			ő		117	1111	60			1 29756	27900	13250	2950	4	4	2	ő	ò	0 Luminasity				1500	24950	ů.		0 1650	6 ē	24950
de.m	5	56956	85.7 Natus VCT	Counter-T Normal				ō					120	1057	141			1 33500	26600	30000	25700	9					1 Luminosity				3500	900	0		0 350	0 0	900
de_m		67930	109 Natus VCT	Terrorists Normal									139	1248	138			0 34950	30750	27500	25900						1 Luminosity				7450	4850			0 745	0 0	4850
de m		81858	82.4 Lumino T	Unknown Normal									130					0 33400	30350	20600							1 Luminosity				2800	9650			0 -1290		+9550
de_m		92413	152 Lumino T	Unknown Normal									0 167	1483	187			0 18500	19200	15850	21500						1 Luminosity				-360	-2300			1 35	0 1	2300
de_m	9	111905	120 Lumino T	Unknown Eco CT	Natus N	6							0 105	992	56			0 11000	1600	20900	21850	8					0 Luminosity			0 -	9900	-20250			1 990	0 1	20250
de_m	10	127294	198 Natus VCT	Terrorists Normal							0		97.4	889	85		0	0 21900	22250	22850	25200	9			0		0 Luminosity			0	960	2950	0		1 -95	0 1	-2950
oe.m		144742	98.4 Natus VCT	Terrorists Force b T	Lumino			1					3 142	1296	110			0 17700	27900	17550	15350	8					3 Luminosity				150	12550			0 15	2 2	12550
20.0	12	107347	ADE NATION OF	Terrorists Porce o T	Luccino		- 2						97.9	910	26		2	0 21000	16800	13300	10500			-	0		0 commonly				1000	11500			0 000		10800
de m	14	170456	93.7 Natura VCT	Terrorists Normal	2011110	ž					0		104	945	95			0 32200	31300	24200	24800	10			0	4	3 Luminasity				1500	6500	ň		0 750	ő ő	6500
de m	15	191445	197 Lumino T	Unknown Force b T	Lumino					0	0		0 163	973	158			0 37700	\$3350	15800	18750	2	8		0		2 Luminosity			- i i	1900	15200		0	0 -1820	6 6	-15200
dem	16	216807	150 Lumino CT	Counter-T Pistol round									120	1110	92			1 4000	4500	4000	4450						0 tuminosity					-150				0 1	150
deum		231859	98.8 Lumino CT	Terrorists Force b CT	turnino								5 85.3					0 19856	11600	12900	22200						0 Luminosity				2210	-10600			0 -695	0 1	106/00
deum		263512	85.4 Lumino CT	Terrorists Doo CT	turnino								26.9	103				0 10310	1600	11000	26120						1 Luminosity				7320	-24550			0 -735	0 1	24950
de_m	19	253641	125 Lumino CT	Terrorists Normal									5 110	972	122			0 30900	22450	22000	31120	11					3 Luminosity				2900	-8700			0 -890	0 1	8700
de_m	20	275467	91.6 Lumino CT	Counter-Titoo CT	Lumino	÷.	4			2	2		5 115	1110	42	°		1 36400	6450	15650	33200	6			2	0	2 Luminosity	0			2750	-26750			0 -2275	9	26750
00_m	21	207220	162 Lumino Ci	Terrorazs Normal			2				0		100	1243	120			0 35650	20100	30/00	34300	15			0		2 cuminatiny				0000	-7400			0 -020		7450
	21	140055	114 Jumino CT	Terroclete Normal							0		60.7	102	70			0 34100	22250	18500	10750	10	6		0		2 Luminosity			0 1	1000	-7400			0 -1560	í i	7400
de m	24	354650	99.9 Lumino CT	Terrorists Eco CT	Lumino			0			0	0	0.60	663	18		0	0 21700	1100	9500	22200	4				0	1 Luminosity	0		0 1	2200	-21100			0 -1220	0 1	21100
de m	25	367433	96 Lumino CT	Terrorists Force b CT	Lumino								2 126	1122	133			0 29450	17600	18400	33600						2 Luminosity				1050	-16000			0 -1105	0 1	16000
de_m		379600	115 Lumino CT	Terrorists Eco CT	Lumino													0 37750	11700	14500	33400						1 Luminosity				3250	-21700			0 -2325		21700
deum		394259	129 Natus VT	Unknown Force b CT	Lumino													0 35850	19300	18500	32500						1 Luminosity				7350	-13200			0 1735	0 1	-13200
de_m		410728	168 Natus VT	Unknown Eco														0 19100		17850							0 Luminosity				1250				0 125	0 1	0
deum	29	432235	173 Natus VT	Unknown Normal		8					0	0	3 124	1106	131			0 12900	22760	1350	21600		6			0	0 Luminosity			1 1	1560	1160	0		0 1165	0 0	1150
de,m	30	454405	160 Natus VT	Unknown Eco T	Natusiv						0		122	1130	59			0 10850	22450	19250	1700	4	3			0	0 Luminosity		9	1 -	3400	20750	0		1 -840	0 0	20750
28.m	91	606931	145 National T	Linksown, Normal		6					0	-	129	1058	117			0 20200	27150	20050	20050	10		-			2 Luminosity		0		-160	-9550			1 .90		-9550
28.m	39	506621	28 Lumino CT	Terrociste Normal							0		159	017	199			0 20200	20000	44950	30550	15	2		0		2 comosity		0		1900	400			1 1895	6 Å	10580
de m	34	536825	290 Natura VT	Linknown Normal					0	0	0	3	3 192	1331	186			0 18050	29600	31550	28000	16	a				3 Luminasity		0		1500	3600		0	1 1350	6 6	3600
de m	35	573973	95.4 Lumino T	Unknown Normal									73.6	427	44			0 10900	27700	21150	26350						0 Luminosity			1 -1	0250	750			1 1025	0 0	-750
de m	36	586191	53.8 Lumino T	Unknown Normal									3 111	280	129			0 30300	28750	42550	27950						0 Luminosity			1 -1	2250	800			1 1225	0 0	-800
de_m		500184	137 Lumino T	Unknown Normal									0 161	1469				0 22850	23300	50900	28350						0 Luminosity			0 -2	5050	-5050			1 2805	0 1	5050
de_m		599164	157 Lumino T	Unknown Normal										1469				0 22850	23300	50900	28350						0 Luminosity			0 -2	5050	-5050			1 2805		5050
de_du		1800	91.5 Natus VT	Unknown Pistol round									128	1169				0 4000	4400	4000	4350						O Natua Vince										-50
de_dt		15507	dool Natur VT	Unknown Eco CT	Astrals								5 61.4	611				0 8500	5650	116/10	16100						O Natua Vince			0 -1	0109	101420			1 1000	3 1	00422
degd.		20105	104 Million T	enerown Eco CT	A10111	-	-							1054				0 12650	1105	1,010	1 100		-	-			O Matcha Vince	0	0		200	-10100			610		MP-00
			and the second se										-	_				-	_	- CONTRACTOR 101							and the second se			-					1010		

MLG 2016 flat file with additional calculated columns

This provides a strong basis to start analyzing various aspects of the event as well as ascertain important levels of performance.

Massaging the Data

The most important column to add to the table is also the most tedious. Before the CSGO-Demos-Manager created column number ¹ it is best to a new column, demo ID. This creates a faux primary key² to compare individual matches with ending match results. Without this, there is no way to pull data related to the overall match from the general sheet to the rounds sheet. To do this, manually copy the relevant Demo ID from column B on the general sheet.

General Sheet

Sadly, CSGO-Demos-Manager <u>improperly symbolicates</u> the winner column. To resolve this, replace that column with the following formula³: =IF(N2>02,L2,M2).

This checks if Team 1's score is larger than Team 2's score. If this case is met, the formula outputs Team 1's name; if the case is not met, it outputs Team 2's name.

Rounds Sheet

The rounds tab can now be related to the general tab thanks to the demo ID column added earlier.

Match Winner

This column references the general sheet and tells us what the winner of the entire match is. This is important when looking for variables that affect the outcome of entire matches and not individual rounds. To do this, use =VLOOKUP(A2,General!B:AT,19,FALSE).⁴

As a sanity check, the match winner column should remain the same for a given set of rounds. It does, so this calculation appears accurate.

Round Winner = Match Winner?

Now that the data shows who won the overall match, it is trivial to see how often a match winner was a round winner. The simple formula =IF(AJ2=E2,1,0) checks if the winner of the round is the same as the winner of the match. It returns a dummy variable, **0**, if the teams do not match and **1** if they do.

As a sanity check, this column should have a larger sum than the number of rows since winners naturally have more round wins. The spreadsheet has 1,112 rows, 663 of which read **1**, so this calculation appears accurate.

ANALYSIS

Now that the data has been adequately prepared, the real work can begin. Both Excel and Stata prove to be integral tools for analyzing and visualizing the data sets.

Excel Analysis

For simpler analysis and visualization, Excel's tools are easy to use and works well.

Pistol Round Impact

The original research question posited that the pistol round positively influenced a team's chance to win the match. This is easily ascertained with a basic Excel analysis.

First, the relevant data must be extracted from the rounds sheet. We can use the match winner column created earlier to make a simple dummy variable to simplify counting. Insert a new column to the right of the dataset called round winner = match winner? and use =IF(AI2=E2,1,0) to check if the Match Winner is the same as the Round Winner. If they are the same, the equation returns **1**. If not it returns **0**.⁵

Once the round winner = match winner? column has been created, it is simple to count the data with a simple =COUNTIF formula. Create two cells far off to the right of the data and use the following formula: =COUNTIFS(H:H, "Pistol round", A0:A0,1). This checks the type column and restricts it to pistol rounds and then counts the frequency of 1 in the round winner = match winner? column. Directly this we use `=COUNTIFS(H:H, "Pistol round", A0:A0,0) to get the frequency of 0 in the same column.

ВА	BB

Match	=COUNTIFS(H:H,"Pistol round",AO:AO,1)
Doesn't Match	=COUNTIFS(H:H,"Pistol round",AO:AO,0)

Simple division with =BB2/COUNTIF(H:H, "Pistol round") and =BB3/COUNTIF(H:H, "Pistol round") allow us to calculate the proportion of wins and losses. According to the data, this results in:

Туре	Value	Proportion
Match	55	67%
Doesn't Match	27	33%



The data shows a 67% match rate, which indicates that the Pistol Round winner is the Match Winner 67% of the time. This is to say that winning the pistol round achieves a 2/3 probability of winning the entire match. To clarify, the word "Match" in the chart indicates rounds where the pistol round winner matches the match winner.

Economic Resource Impact

Before calculating the impact resources has on a match, the data must describe the frequency of economic impact on individual rounds. This can partially be calculated in Excel, however more advanced analysis will require Stata.

First, it is necessary to create a new formula that outputs a dummy variable. **1** should define when the winner has more resources, **0** should define when the winner has less. The simplest way to accomplish this is to create a new column called winner spend more on equipment? and reference the general sheet and match winner column:

=IF(IF(Z2>AB2,VL00KUP(A2,General!B:AT,11,FALSE),VL00KUP(A2,General!B:A T,12,FALSE))=AI2,1,0).

This formula uses a nested =IF. The interior statement checks if Team 1 starts with more equipment than Team 2. If Team one starts with more, the formula pulls Team 1's name from the general sheet. If Team 2 starts with more, it pulls that name instead. Next, the exterior statement compares the string result from the interior statement to the string in the match winner column. If the strings match, Team 1 is the winner, and the formula outputs a **1**. If Team 2 is the winner, the formula outputs **0**. This column now calculates which team has better equipment and if having better equipment corresponds with a win.

BA

BB

Winner has More Equipment =COUNTIF(AR:AR,1) Winner has Less Equipment =COUNTIF(AR:AR,0)

As before, use =COUNTIF(AR:AR,1) and =COUNTIF(AR:AR,0) to count the frequency at which this specific test occurs. This formula is far simpler than before because we are not using other columns to restrict counting.

Туре	Value	Proportion
Winner has More Equipment	633	60%
Winner has Less Equipment	449	40%



Further, =BB8/SUM(BB8:BB9) and =BB9/SUM(BB8:BB9) can be used to calculate the relevant proportion.

The winner has better equipment 60% of the time — barely more than half. This demonstrates that, at the professional level, the game's economy has little effect on various outcomes.

Prepping for Stata

Sadly, Excel hits a wall when it comes to more advanced econometric plotting and analysis. However, Stata's data entry system can be slow and arduous; it is often a better choice to use Excel to create the data and then simply copy it into Stata. This section will describe several new columns that will not be used in the Excel analysis but will come into play for the Stata section.

AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT
Match Win	ls Team 1	Does Tean	Does Tean	Start Tean	Spend Tea	Round wir	CT or T?	Winner Sta	Winner Sp	Difference	Difference
Luminosity	1	I 0	1	0	100	1	1	0	1	I 0	100
Luminosity	1	I 1	1	10250	10800	1	1	1	1	I 10250	10800
Luminosity	1	I 1	1	10900	20200	1	1	1	1	I 10900	20200
Luminosity	1	I 1	1	12850	10800	1	1	1	1	I 12850	10800
Luminosity	1	I 1	1	16150	9800	1	1	1	1	I 16150	9800
Luminosity	1	I 1	1	19500	13450	1	1	1	1	I 19500	13450
Luminosity	1	I 1	1	12100	24150	1	1	1	1	I 12100	24150
Luminosity	C) 1	1	7750	6800	0	0	1	1	I -7750	-6800
Luminosity	C) 1	1	12000	4400	0	0	1	1	I -12000	-4400
Luminosity	1	I 0	1	-8400	5450	1	1	0	1	I -8400	5450
Luminosity	1	I 0	1	-2300	3400	1	1	0	1	-2300	3400
Luminosity	1	I 1	1	8300	25550	1	1	1	1	I 8300	25550
Luminosity	1	ı 1	1	8350	8900	1	1	1	1	8350	8900

These columns in the rounds sheet are as follows:

- Match Winner
 - Outlined above
 - =VLOOKUP(A2,General!B:AT,19,FALSE)
- Is Team 1 the Winner?
 - Checks if Team 1 is the winner based on the Team 1 in the general sheet.
 - Dummy variable that returns 1 if Team 1 is the winner, 0 if not
 - =IF(VLOOKUP(A2,General!B:AT,11,FALSE)=E2,1,0)
- Does Team 1 Have More Money?
 - Checks to see if Team 1 has more money than Team 2.
 - Dummy variable that returns 1 if Team 1 has more money, 0 if not
 - =IF(Y2>AA2,1,0)
- Does Team 1 have More Equipment?
 - Dummy variable that returns **1** if Team 1 has better equipment, Oso this calculation appears accurate if not
 - =IF(Z2>AB2,1,0)
- Start Money Team 1 Start Money Team 2
 - Determines the difference between Team 1's economy and Team 2's economy
 - =Y2-AA2
- Spend Team 1 Spend Team 2⁶
 - Determines the difference between Team 1's buy and Team 2's buy
 - =Z2-AB2
- Round Winner = Match Winner?
 - Outlined above
 - =IF(AI2=E2,1,0)
- CT or T?
 - · Checks if the winning team is counter-terrorist or terrorist

- Returns a dummy variable 1 if the winning team is counter-terrorist and 0 if not
- =IF(F2="CT",1,0)
- Winner Start with More Money?
 - Uses a nested =IF to create a dummy variable 1 if the Winner starts with more money and **0** if not
 - Similar to winner spend more on equipment outlined above
 - =IF(IF(Y2>AA2,VL00KUP(A2,General!B:AT,11,FALSE),
 VL00KUP(A2,General!B:AT,12,FALSE))=AI2,1,0)
- Winner Spend More On Equipment?⁷
 - Outlined above
 - =IF(IF(Z2>AB2,VL00KUP(A2,General!B:AT,11,FALSE),
 VL00KUP(A2,General!B:AT,12,FALSE))=AI2,1,0)
- Difference Between Winner and Loser Starts
 - Calculates the difference in resources between Team 1 and Team 2.
 - =IF(VLOOKUP(A2,General!B:AT,11,FALSE)=E2, Y2-AA2,AA2-Y2)
- Difference Between Winner and Loser Spends⁸
 - Calculates the difference in equipment value between Team 1 and Team 2.
 - =IF(VL00KUP(A2,General!B:AT,11,FALSE)=E2, Z2-AB2,AB2-Z2)

Stata Analysis

State provides a robust platform for advanced econometric analysis.

Charting Kill Frequency

The first visualization in Stata is relatively simple. First, we must adjust the kill data on the players sheet of *MLG2016Full.xlsx*. This adjustment will account for the amount of rounds players participated in. If the data is not adjusted, players that played more rounds that have higher overall totals will end up above players that have a higher ratio of kills per round.

To accomplish this, add 5 columns, one for each kill quantity.⁹ Because the number of rounds played is static, the formula =L2/\$J2 can be used to calculate the correct ratios. Locking the denominator to the J column ensures the rounds column is always used while also allowing the formula to be filled relatively.

Next, the kill totals must also be adjusted. Adding a column next to kills with =D2/C2 ¹⁰ will adjust the kill count to a per-match ratio. If we do not adjust the kill totals, we get skewed output like this:



Total kills plotted against adjusted kill quantity

	L	М			Ν			0			Ρ		Q		R			S	т		U	I
5K	-	4K	-	ЗК		•	2K		▼	1K		▼	Adj5	▼	Adj4	▼	Adj3	-	Adj2	▼	Adj1	-
	1		0			6			20			60	0.004	195		0	0	.0297	0.09	901	0.29	9703
	0		0			4			19			49		0		0	0.0	2685	0.12	2752	0.32	2886
	0		1			5			21			60		0	0.00	546	0.0)2732	0.11	475	0.3	2787
	0		1			1			12			33		0	0.01	053	0.0	01053	0.12	632	0.3	4737

0.0146

0.01266

0.0038

0.16058

0.09836

0.1407

0.1519

0.11027

0.06818

0.1358

0.0219

0.01639

0.03015

0.03165

0.04943

0.04321

0.25547

0.26776

0.27638

0.3038

0.26996

0.31818

0.24074

The file *MLGPlayerData.dta* attached below already has these columns inserted.

Location of unadjusted and adjusted data columns

Once the data is imported to Stata, we use the scatter command to make the scatterplot. To scatter Total Kills to the Adjusted 1K Ratio, enter scatter killspermatch adj1, mlabel(name). To scatter the other adjusted kill ratios, use adj2, adj3, and so on. Adding ,mlabel(name) tags each point with the corresponding player name.



Adjusted total kills plotted against adjusted kill quantity

Vertical height denotes more overall kills per round. Horizontal extension denotes the propensity of a player to hit that certain amount of kills: the further right a player is, the more often they hit that number of kills per round.

I omitted adj5 from this visualization because only 7 players achieved 5k.¹¹ There is not enough data to draw a reasonable conclusion.

Equipment Value

Each round rewards players with various income based on individual and group performance. Regression models can describe how the gameplay economy affects match outcomes.

The rounds sheet holds all of the data used for the regression analysis. Thanks to the columns added prior, copying and pasting directly into Stata allows us to dove right into the data.¹²

These imported variable names are quite long. Luckily, Stata has an autocomplete feature: press the **Tab** key to bring up a list of variables that complete the text entered in the command line.



Regressing variables onto isteam1thewinner will find the impact said variables have on influencing Team 1 to beat Team 2. For example, reg isteam1thewinner equipementvalueteam1 equipementvalueteam2 does a multiple regression to ascertain the effects variables equipementvalueteam1 and equipementvalueteam2 have on the variable isteam1thewinner. This command results in:

Source	5	SS o	lf MS	Num	ber of obs		1,1	12
Model Residual	60.904 216.39	40756 90888 1,10	2 30.45203 09 .1951225	378 Pro 32 R-s	$rac{1109}{b > F}$ quared	=	0.00	07 00 96 82
Total	277.29	94964 1,12	.2495904	127 Roo	t MSE	=	.441	73
isteam1the	ewinner	Coef.	Std. Err.	t	P> t	[95%	Conf.	Interval]
equipementvalı equipementvalı	ueteam1 ueteam2 _cons	.0000164 0000186 .5247827	1.30e-06 1.37e-06 .0389365	12.61 -13.59 13.48	0.000 0.000 0.000	.0000 0000 .4483	0138 0213 3852	.0000189 0000159 .6011802

Regressing variables equipementvalueteam1 and equipementvalueteam2 onto isteam1thewinner.

This tells is that β_2 and β_3 — the respective effects of equipementvalueteam1 and equipementvalueteam2 — have inverse impacts on Team 1 winning a round (y). β_1 , the constant, shows that if both teams are equal, there is a 50% chance that team 1 wins.

The full equation reads y = 0.0000164x - 0.0000186z + 52.5 where x is the dollar value of Team 1's gear and z is the dollar value of Team 2's gear. Each dollar of Team 1 gear raises their chance of winning by 0.00164% and each dollar of Team 2 Gear lowers Team 1's chance of winning by 0.00186%. All things considered, this results in very little economic impact on professional gameplay unless there are large differences in the buying power of each team.

The variable doesteam1havemoreequipment allows us to perform another simple regression onto isteam1thewinner, this time to determine the overall effect of having more equipment.

Source	SS	df	MS	Number of o	bs =	1,112	
				F(1, 1110)		128.96	
Mode I	28.8633485	1	28.8633485	Prob > F		0.0000	
Residual	248.431615	1,110	.223812266	R-squared		0.1041	
				Adj R-squar	ed =	0.1033	
Total	277.294964	1,111	.249590427	Root MSE		.47309	
iste	am1thewinner	Coe	f. Std. Ern	•. t	P> t	[95% Conf.	Interval]
doesteam1havem	oreequipment	.32222	36 .0283743	11.36	0.000	.2665501	.377897
	_cons	.31283	.0201177	15.55	0.000	.273366	.3523122

Regressing variable doesteam1havemoremoney onto isteam1thewinner

The constant here is approximately 0.31 while β_2 is about 0.32. This means the regression line would be y=0.32x + 0.31 where y is the percent chance of winning. This tells us that there is a 31% chance of winning when Team 1 has less equipment and a 63% chance of winning when Team 1 has more equipment. This confirms the earlier Excel analysis that posited the winner has more equipment 60% of the time:

Туре	Value	Proportion
Winner has More Equipment	633	60%
Winner has Less Equipment	449	40%

Eco and Semi-Eco

Value of team inventories occasionally forces teams to save and purposely not buy inventory — a strategy dubbed *Eco* or *Semi-eco* to try and push back later. CSGO-Demos-Manager is able to automatically analyze the type of round and conveniently creates a type column that contains the information:

. tabulate type			
Туре	Freq.	Percent	Cum.
Eco	344	30.94	30.94
Force buy	55	4.95	35.88
Normal	588	52.88	88.76
Pistol round	82	7.37	96.13
Semi-Eco	43	3.87	100.00
Total	1,112	100.00	

To study the effects of this strategy the data needs to be restricted to only show Eco and Semi-eco rounds: simply filter the type column on the round sheet.¹³



Once the data has been restricted and imported to Stata, regressing doesteam1havemoremoney onto isteam1thewinner elucidates the effects of better equipment:

. reg isteam1thewinner doesteam1havemoreequipment							
Source	SS	df	MS	Number of obs		387	
Model Residual	61.0242475 35.6165793	1 61 385 .0	0242475	F(1, 385) Prob > F R-squared		659.65 0.0000 0.6315	
Total	96.6408269	386 .2	50364836	Adj R-squared Root MSE		.30416	
isteam1thewinner		Coef.	Std. Err.	t P>	t	[95% Conf.	Interval]
doesteam1havemoreequipment _cons		.7986873 .1261682	.0310972	25.68 0. 6.07 0.	000 000	.7375457 .0852888	.8598289 .1670476

The regression line is y = 0.79x + 0.126. This tells us that having more equipment $-\beta_2$ – raises the odds of winning an eco or semi-eco round by 79%. While this is high, it means that variables other than equipment affect the outcome of eco rounds – but only 9% of the time.

RAW DATA

This data comes in 3 major formats: Demo (.dem), Excel (.xlsx), and Stata (.dta). HLTV.org hosts the demo files. All of the files used are located in the <u>/files</u> directory of this post. These are the individual files with respective descriptions:

- MLG2016Full.xlsx (461 kb) | Direct Download | View Google Sheet
 - Annotated and formatted like this article describes
 - Excel file with 6 sheets of data summarizing the entire event:
 - general contains a summary for each match
 - winner column augmented with formula =IF(N2>02,L2,M2)
 - players contains statistics for each player
 - K/D column augmented with formula =D2/F2
 - maps contains map statistics
 - teams contains team statistics
 - weapons contains various weapon statistics
 - Contains additional calculated columns
 - damage per shot calculated with =D2/E2
 - damage per hit calculated with =C2/F2
 - rounds contains round statistics
 - Contains several formulaic columns separated from static columns by a vertical black line; index included beside the data.
- *MLGTeamData.dta* (26 kb) | <u>Direct Download</u>
 - Stata file with variables pulled from the teams sheet of MLG2016Full.xlsx
- *MLGPlayerData.dta* (45 kb) | <u>Direct Download</u>
 - Stata file with variables pulled from the players sheet of *MLG2016Full.xlsx*
- *MLGFull.dta* (256 kb) | <u>Direct Download</u>
 - Stata file with variables pulled from the rounds sheet of *MLG2016Full.xlsx*
- *rawData.zip* (1.1 mb) | <u>Direct Download</u>

- Contains an unannotated copy of *MLG2016Full.xlsx* as well as individual data exports for each match in the Columbus 2016 competition folder according to level of bracket.
- This is the raw data that CSGO-Demos-Manager spits out after analyzing demo files and has not been organized.

DISCLOSURE AND NOTES

- All data was sourced from HLTV-hosted match demos
- Findings, calculations, and all data are published under the CC-BY-4.0 license
- EnVyUs v. CLG match demo was corrupted and thus the data was not accessible

Software Used

Proprietary

- Microsoft Excel
 - Data organization, calculation, and visualization
- StataCorp <u>Stata</u>
 - Regression analysis and visualization

Open Source

- <u>CSGO-Demos-Manager</u>
 - Demo parsing

Discussion on <u>r/GlobalOffensive</u>

- 1. Originally "A"
- 2. Although it is not technically unique, this value serves to relate separate pieces of the data.

- 3. This assumes you have not modified the existing order of the columns. If you have, use an if formula to check if Team 1's score is larger than Team 2's score. If true, output Team 1's name; if false, output Team 2's name.
- 4. Again, if you have moved cells around, you will need to use different references.
- 5. One could also accomplish this without creating a new column by using =COUNTIFS or =SUMIFS and focussing on all three relevant columns, however because later analysis uses the round winner = match winner? column it makes sense to create it now.
- 6. Actually "Equipment Value" and not "Spend" per se.
- 7. Actually "Has Better Equipment" and not "Spend More On" per se.
- 8. Actually "Equipment Values" and not "Spends" per se.
- 9. Kill quantity refers to how many kills a single player gets in a given round, i.e. 1k is one kill, 2k is 2 kills, 3k is three kills, and so on.
- 10. Inserting a column here will preserve the references to the right.
- 11. As a result, everyone is bunched up above the X axis at 0.
- 12. *MLGFull.dta* located below contains this data in a pre-made Stata file.
- 13. Another solution is to <u>use</u> gen or tabulate in Stata, however it is often simpler to adjust the data using Excel filters and create a new Stata file.