Competitive Esports by the Numbers

Identifying sources of team success and effective individual player performance in Counter Strike Global Offensive

AUTHOR1 and AUTHOR2

CSE 163
Hunter Schafer
17 May 2019

PART 2

Research Questions:

We will be focusing on four major questions:

1. What makes a successful team? What statistics do most teams that rank high in tournament placements share? In winning teams, do players all participate equally in getting objectives and kills? Do ethnically homogenous teams perform better than those that aren't? Is it better to be a good team on certain maps?
    a. **Answer:** Higher rated teams seems to have a higher mean ADR (Average damage per round), assists, KAST (Percentage of rounds where a player either gets a kill, an assist, survives the round, or is responsible for another enemy killed via death by distraction), kills per game, and opening kills per game. Lower rated teams tend to have more deaths and opening deaths per game. Hit flashbangs were the only statistics that didn't seem to rise when looking at better teams. For most teams, players all contribute about equally to KAST, and equally to deaths (or the lack thereof). Most teams have a larger distribution in player performance when it comes to opening kills and kills in general, as well as successful flashbangs. ADR and assists are little more consistent, but still fluctuate a lot between players on a team. In terms of whether or not ethnically homogenous teams (At least 3 players are the from the same country/region) are better than "international team", we found that international teams fell behind in hit flashbangs and assists, stayed on par in ADR, kills, opening kills, and KAST, and actually led in headshots and deaths/opening deaths. Most high rated teams maintain high win rates on the maps Overpass, Train, and Inferno, Dust2 and Nuke are less popular. Some teams, like Astralis probably have their spot in the top 30 due in some part to high win rates on specific maps (~90% on Nuke).
2. What makes a successful player? What common traits do successful players share? What statistics factor in to a player getting a spot on the top 20 list each year?
    a. **Answer**: The most highly ranked players have the most kills/opening kills. Everyone in the top 20 has a consistently high number of kills every match and average damage per round. Assists and headshots are not as important factors, but some players on the top 20 are have significantly higher stats in those categories than others.
3. What countries/regions produce the best teams and players? What countries have the highest representation of teams and players? What regions have the largest percentages of highly ranked teams and players?
    a. **Answer:** NA, Scandinavia, and Northwestern Asia produce the strongest players (i.e. Denmark, Sweden, Canada, Russia). Denmark is ahead by a large margin when it comes to representation in the top 30 teams. The US, Sweden, and Brazil come in as close seconds.
4. How much is "Home-Field advantage" a factor? Do teams performing in their home regions tend to do better than when in tournaments away from home?
    a. **Answer:** There is no such things as home field advantage. Cloud9 exceeded all expectations at Boston 2018, ENCE and TYLOO play consistently better in their home countries. But MIBR and Renegades seem to play the same if not worse in their home countries.

Motivation and background: We ask these questions to look for trends in gameplay from CS:GO. Professional CS:GO is a high-stakes game and plays well for both players and organizations. Finding correlation between certain variables and success is valuable information and is interesting to examine the nature of human competition. Knowing what variables are more closely aligned with winning would shed some light on the best strategies for fast-paced, co-op problem solving.

Data:

- Scraped CSV data fro HLTV.org can be found [here]
- Data for Country Geometry is taken from https://datahub.io/core/geo-countries
- We will be using HLTV.org's top 30 list as a base for determining how "successful" teams are and
- Their top 20 players of 2018 list, similarly for players

Matches.csv contains information about single CSGO matches, including the two teams playing, the winner, date, tournament, players present, and other stats such as the highest rated players in the match, who did the most damage, got the most AWP kills, etc. Looks something like this but much wider:

matchId	tournamer	date	map	team1	team2	winner	team1_rounds	team2_rounds	team1_first_half	team1_second_half	team2_first_half	team2_second_half
85760	BLAST Pro	5/11/2019 9:15	Train	Astralis	ENCE	ENCE	13	16	8	5	7	9
85758	BLAST Pro	5/11/2019 9:15	Nuke	Astralis	ENCE	ENCE	9	16	4	5	11	5
85748	BLAST Pro	5/11/2019 6:50	Inferno	Astralis	Cloud9	Astralis	16	11	9	7	6	5
85741	BLAST Pro	5/11/2019 5:25	Dust2	Astralis	ENCE	ENCE	13	16	8	5	7	9
85730	BLAST Pro	5/11/2019 4:00	Dust2	Astralis	Natus Vinc	Astralis	16	13	5	11	10	3
85691	BLAST Pro	5/10/2019 11:20	Nuke	Astralis	Giants	Astralis	16	7	9	7	6	1
85684	BLAST Pro	5/10/2019 10:00	Dust2	Astralis	NiP	Astralis	16	8	9	7	6	2
85039	ESL Pro Lea	4/25/2019 11:00	Overpass	Astralis	BIG	Astralis	16	14	5	11	10	4
85034	ESL Pro Lea	4/25/2019 11:00	Inferno	Astralis	BIG	BIG	17	19	9	6	6	9
85032	ESL Pro Lea	4/25/2019 11:00	Dust2	Astralis	BIG	Astralis	16	10	10	6	5	5
84986	ESL Pro Lea	4/24/2019 7:20	Inferno	Astralis	HellRaisers	Astralis	16	7	9	7	6	1
84954	ESL Pro Lea	4/23/2019 11:15	Nuke	Astralis	ex-3DMAX	Astralis	16	7	8	8	7	0
84949	ESL Pro Lea	4/23/2019 11:15	Mirage	Astralis	ex-3DMAX	Astralis	16	6	9	7	6	0
84364	BLAST Pro	4/13/2019 10:05	Dust2	Astralis	FaZe	FaZe	5	16	2	3	13	3
84360	BLAST Pro	4/13/2019 8:30	Overpass	Astralis	Liquid	Liquid	14	16	8	6	7	9
84352	BLAST Pro	4/13/2019 7:00	Overpass	Astralis	MIBR	MIBR	2	16	2	0	13	3
84296	BLAST Pro	4/12/2019 13:30	Inferno	Astralis	Natus Vinc	Astralis	16	7	10	6	5	2

Players.csv contains information about individual player performance per match, things like their name, origin, and their kills, deaths, assists, enemies flashbanged, headshots, and other personal stats. Looks like this:

name	origin	matchId	tournamer	team	against	kills	headshots	assists	hit_flashbangs	deaths	kast	adr	opening_kills	opening_deaths	rating
Magisk	Denmark	85760	BLAST Pro	Astralis	ENCE	26	10	2	0	18	72.40%	91.2	6	4	1.34
device	Denmark	85760	BLAST Pro	Astralis	ENCE	21	3	2	0	17	75.90%	77.4	2	1	1.13
gla1ve	Denmark	85760	BLAST Pro	Astralis	ENCE	20	7	5	2	20	55.20%	76	4	4	1.01
dupreeh	Denmark	85760	BLAST Pro	Astralis	ENCE	11	6	8	0	21	65.50%	62	0	5	0.71
Xyp9x	Denmark	85760	BLAST Pro	Astralis	ENCE	11	3	5	2	19	62.10%	40.1	1	2	0.62
Aerial	Finland	85760	BLAST Pro	ENCE	Astralis	25	12	5	0	19	72.40%	79.7	7	3	1.27
xseveN	Finland	85760	BLAST Pro	ENCE	Astralis	21	8	3	0	20	62.10%	93.8	2	3	1.05
Aleksib	Finland	85760	BLAST Pro	ENCE	Astralis	21	7	5	4	17	62.10%	76.2	1	1	1.04
sergej	Finland	85760	BLAST Pro	ENCE	Astralis	14	5	5	1	17	72.40%	67.5	4	5	0.96
allu	Finland	85760	BLAST Pro	ENCE	Astralis	14	1	5	1	16	62.10%	51.3	2	1	0.83
device	Denmark	85758	BLAST Pro	Astralis	ENCE	18	6	1	0	16	56.00%	69.4	4	4	1.06
Magisk	Denmark	85758	BLAST Pro	Astralis	ENCE	13	3	3	0	18	64.00%	66.4	2	3	0.87
dupreeh	Denmark	85758	BLAST Pro	Astralis	ENCE	11	4	2	0	20	60.00%	66.3	1	2	0.7
gla1ve	Denmark	85758	BLAST Pro	Astralis	ENCE	14	8	9	2	20	64.00%	68.1	3	2	0.91
Xyp9x	Denmark	85758	BLAST Pro	Astralis	ENCE	11	2	0	0	18	68.00%	45.4	1	3	0.68
sergej	Finland	85758	BLAST Pro	ENCE	Astralis	17	4	3	1	10	84.00%	72.8	0	3	1.29
Aerial	Finland	85758	BLAST Pro	ENCE	Astralis	21	4	7	0	15	84.00%	104.8	8	2	1.59

Tournaments.csv contains information about each tournament, including the name of the tournament, the location, prize pool,and number of teams attending. We will use this to append tournament location to the matches dataset.

name	location	prize_pool	teams_attending
BLAST Pro	Spain	$250,000	6
ESL Pro Lea	United Kingdom	TBA	16
BLAST Pro	United States	$250,000	6
BLAST Pro	Brazil	$250,000	6
ECS Season	Europe	$25,000	8
IEM Katow	Poland	$1,000,000	16
iBUYPOWE	United States	$200,000	8
BLAST Pro	Portugal	$250,000	6
ESL Pro Lea	Denmark	$750,000	16
ECS Season	United States	$660,000	8
ESL Pro Lea	Europe	$105,000	14
IEM Chicag	United States	$250,000	16
BLAST Pro	Denmark	$250,000	6
ECS Season	Europe	$45000 + 4 Finals s	10
BLAST Pro	Turkey	$250,000	6
FACEIT Ma	United Kingdom	$1,000,000	16
FACEIT Ma	United Kingdom	8 spots at FACEIT M	16

Countries.geojson contains geometry data of multiple countries around the world. We will use it to map region data and visualize it.

Top30.csv and Top20.csv contain the top 20 players of 2018, and top 30 teams of csgo as of June 3rd 2019. These both contain the ranking of each team and player, and will be used for sorting.

Top30.csv

position	name
1	Liquid
2	Astralis
3	ENCE
4	Vitality
5	FaZe
6	Natus Vincere
7	fnatic
8	MIBR
9	NiP
10	NRG
11	FURIA
12	Renegades
13	G2
14	mousesports

Top20.csv

name	origin	position
s1mple	Ukraine	1
device	Denmark	2
NiKo	Bosnia and	3
electronic	Russia	4
dupreeh	Denmark	5
NAF	Canada	6
Magisk	Denmark	7
gla1ve	Denmark	8
KRIMZ	Sweden	9
coldzera	Brazil	10
GuardiaN	Slovakia	11
Twistzz	Canada	12
Xyp9x	Denmark	13

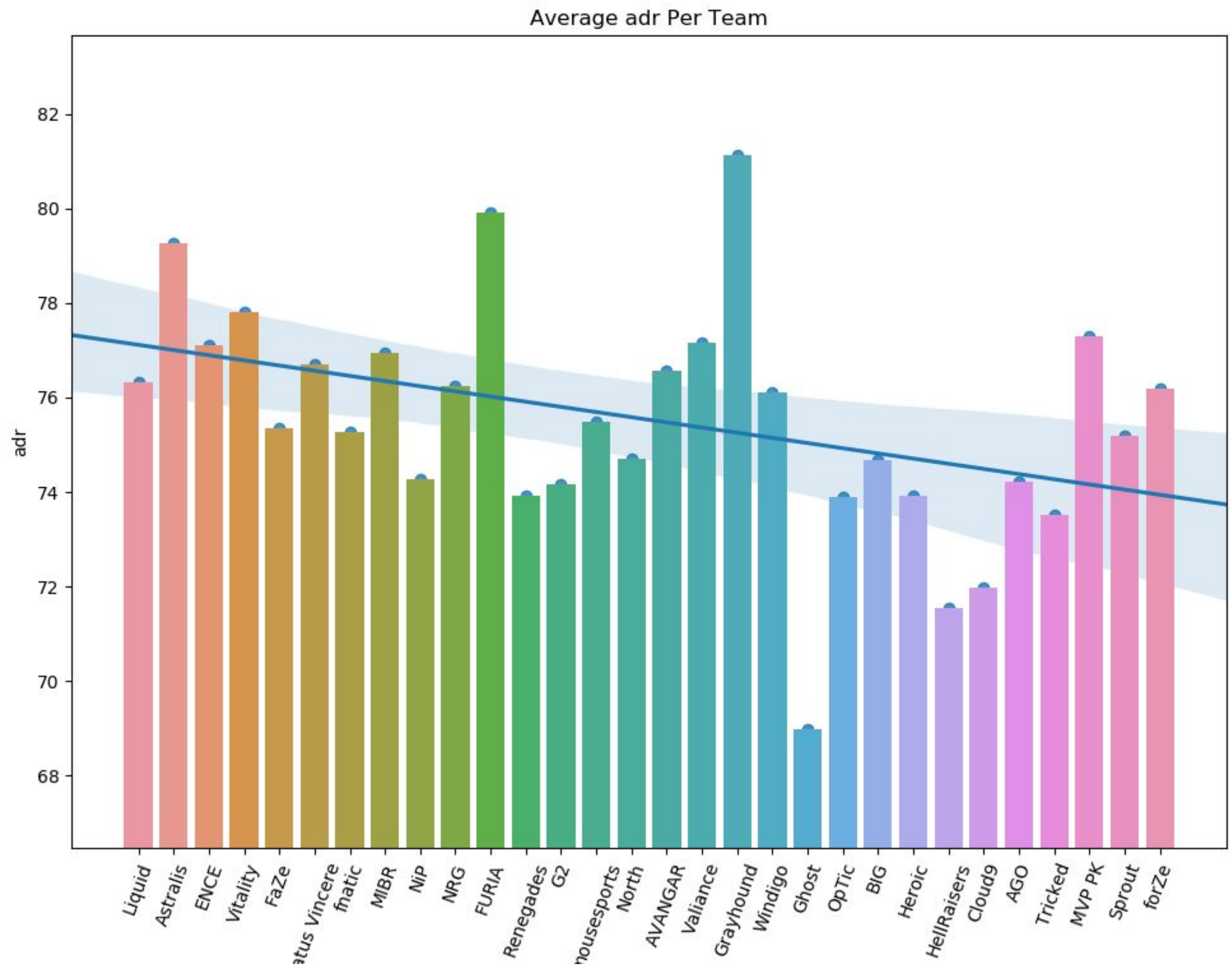
Methodology and <u>Results:</u>

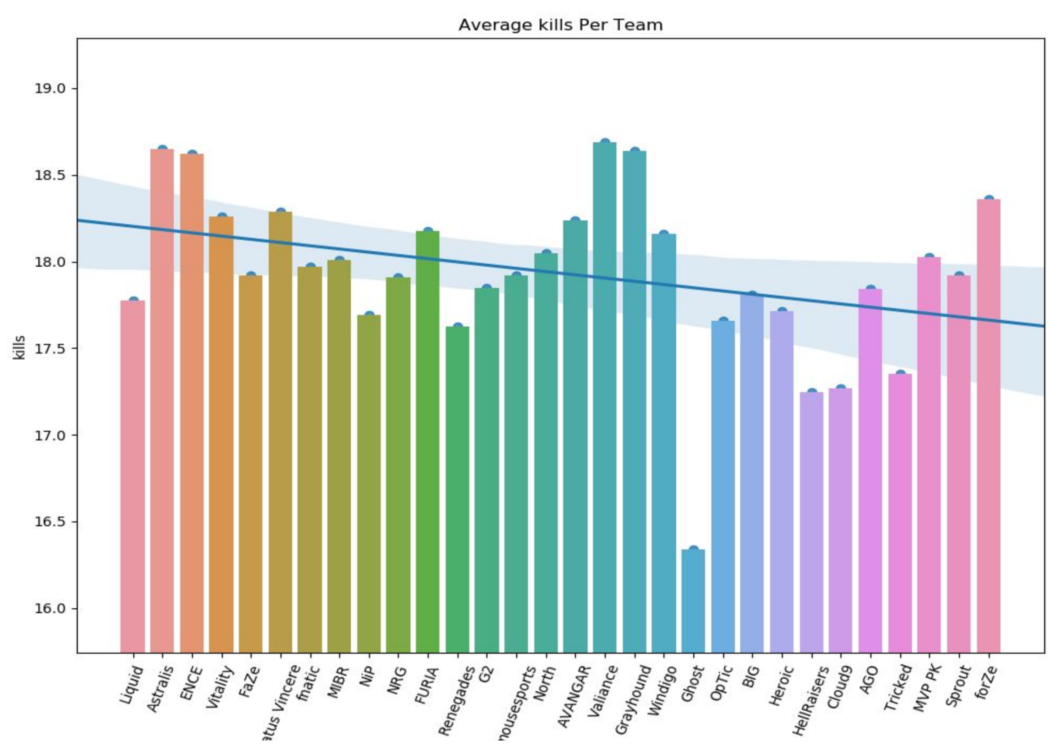
Our research breaks down into 4 different questions.

1. Determining exactly what the recipe for a winning team is isn't easy work, but by examining statistics on a match by match basis, we can easily come to general conclusions about what factors into a victory. Looking at each category suggests what fundamentals are more important- more correlation between assists and wins may suggest team play is important, correlation with opening kills and wins may suggest being quick to the punch and getting an advantage early helps win, and clutches correlating with wins suggests performing well under pressure is most important. By comparing our findings to the HLTV top 30 list, which is representative of tournament wins and therefore actual physical victories, we can find correlations between certain statistics and successful careers. We can easily get a sense of what statistics are higher or lower in winning teams by grouping each team's matches and averaging their statistics. We could do the same for player performance by using the players.csv dataset, grouping and averaging game stats for each player per team, and visualizing how even the spread for kill contribution, utility (flashbang grenades, assists…) usage, and deaths look. From that, we can judge how often teams that have 5 consistent players make it into the top 30, versus teams that have 1-2 "Star players". Since we also have player origins, we visualize how certain teams perform during player lineups without language barriers, versus teams that have players from various different regions. Finally, for team performance, we can examine what maps are played the most in tournaments, and analyze win/loss ratios on those maps per team to get a sense of whether or not having a good ratio on a certain map influences your chance to win.

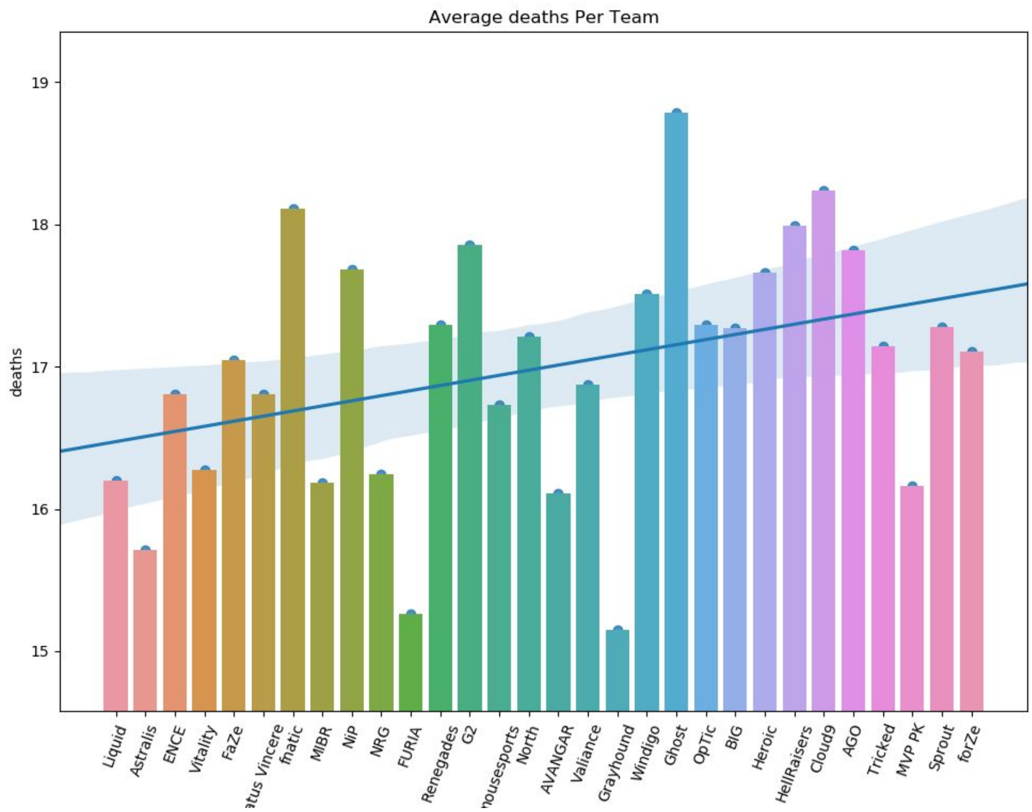
In general, we found that most teams in the top 30 shared a consistently high number for each statistic we looked at. A lot of the graphs tend to visualize small differences, which in a game like CSGO can still mean a lot, especially since some stats like average assists per game are usually smaller numbers anyways.

Teams towards the top end of the top 30 list seemed to have a higher ADR (Average damage per round), KAST (Percentage of rounds where a player either gets a kill, an assist, survives the round, or is responsible for another enemy killed via death by distraction), opening kills, total kills, and number of assists. Here are a few graphs depicting this with the x axis depicting each team, ordered ranking wise left to right, and the y axis representing average of each statistic. The rest can be found in the "stat_per_team" folder.


Average adr Per Team
82
80
78
76
74
72
70
68
adr
Liquid
Astralis
ENCE
Vitality
FaZe
atus Vincere
fnatic
MIBR
NiP
NRG
FURIA
Renegades
G2
nousesports
North
AVANGAR
Valiance
Grayhound
Windigo
Ghost
OpTic
BIG
Heroic
HellRaisers
Cloud9
AGO
Tricked
MVP PK
Sprout
forZe

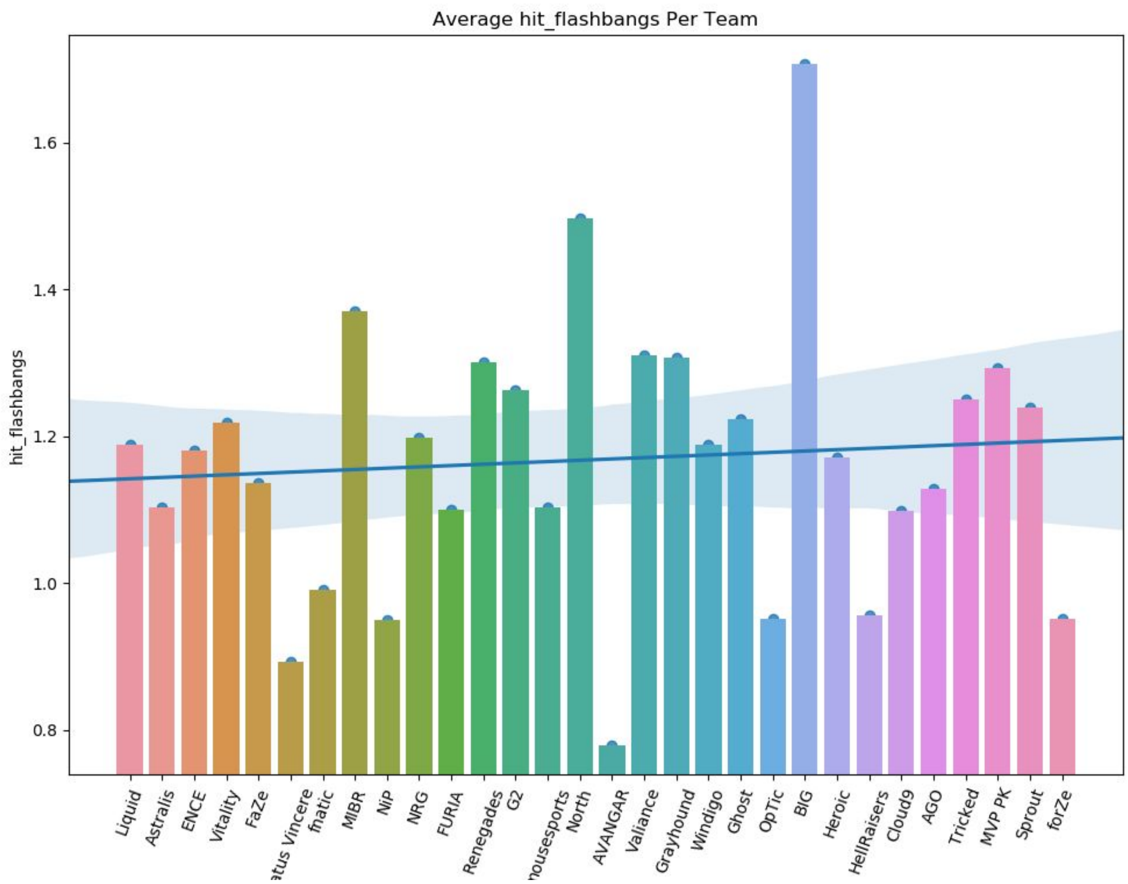

Average kills Per Team
19.0
18.5
18.0
17.5
17.0
16.5
16.0
kills
Liquid
Astralis
ENCE
Vitality
FaZe
atus Vincere
fnatic
MIBR
NiP
NRG
FURIA
Renegades
G2
nousesports
North
AVANGAR
Valiance
Grayhound
Windigo
Ghost
OpTic
BIG
Heroic
HellRaisers
Cloud9
AGO
Tricked
MVP PK
Sprout
forZe

Stats that seemed to be lower for higher rated teams were deaths and opening deaths, as would be expected. Better teams stay alive more often, regardless of whether or not they win the round.

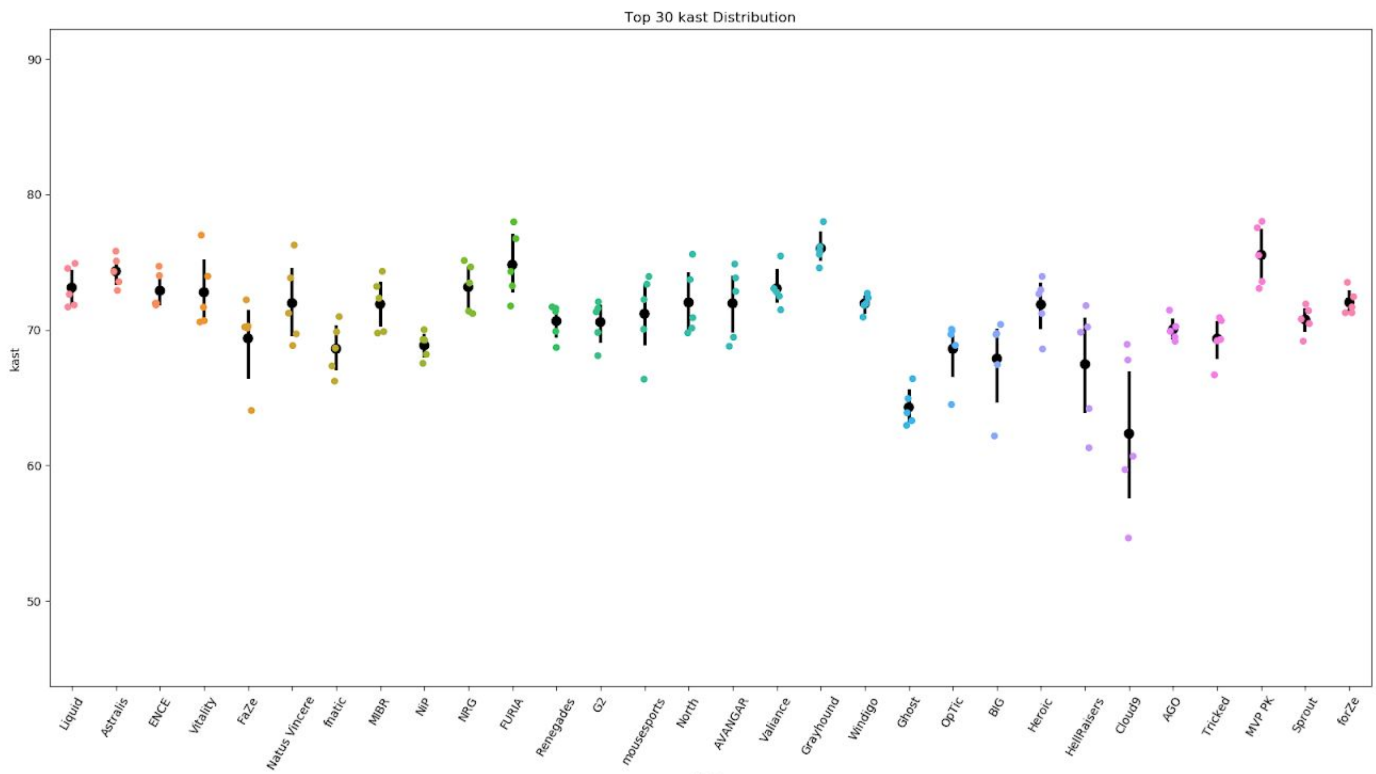


\

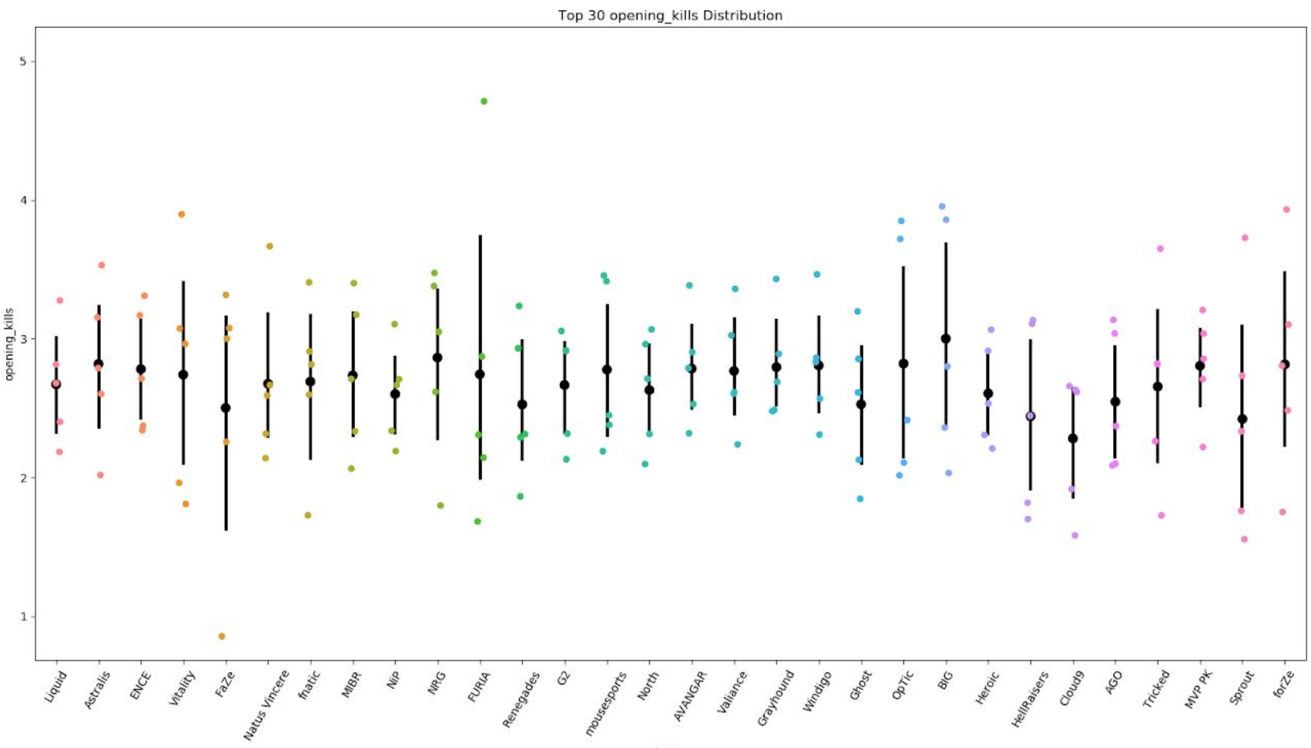
The only statistic that didn't seem to change based on a team's ranking (stayed consistent across all 30) was hit_flashbangs. This could be because teams don't know how to capitalize on flashbangs, so they don't impact the game as much as they should, or maybe because flashbangs can only do so much, so once you learn how to use them as a team, you hit a skill ceiling.



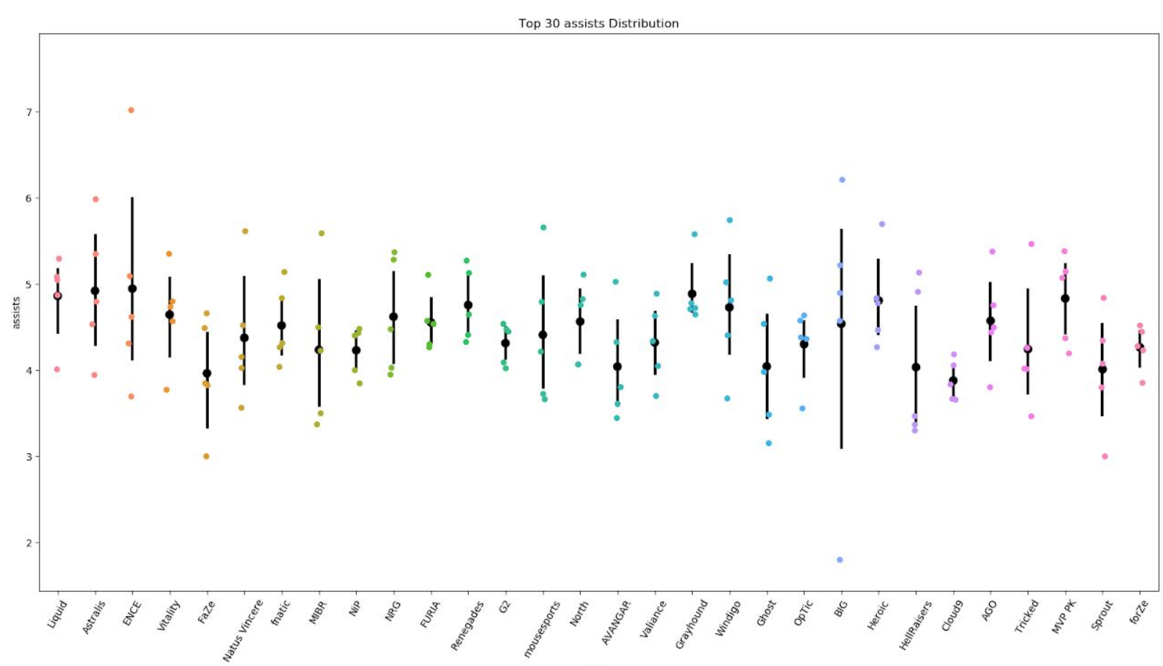
When looking at performance distribution across teams, we found that players across all teams mostly contribute equally in terms of KAST and deaths (or the lack thereof). It makes sense that in high performing teams, all 5 players are required to contribute something positive to the round, and KAST is a statistic that encompasses most positive attributes of a round. You can tell from the graph below how tight each player is to one another, the black lines representing mean distribution being considerably smaller than graphs of other statistics. Some graphs are shown below. The rest are in the “stat_dist” folder



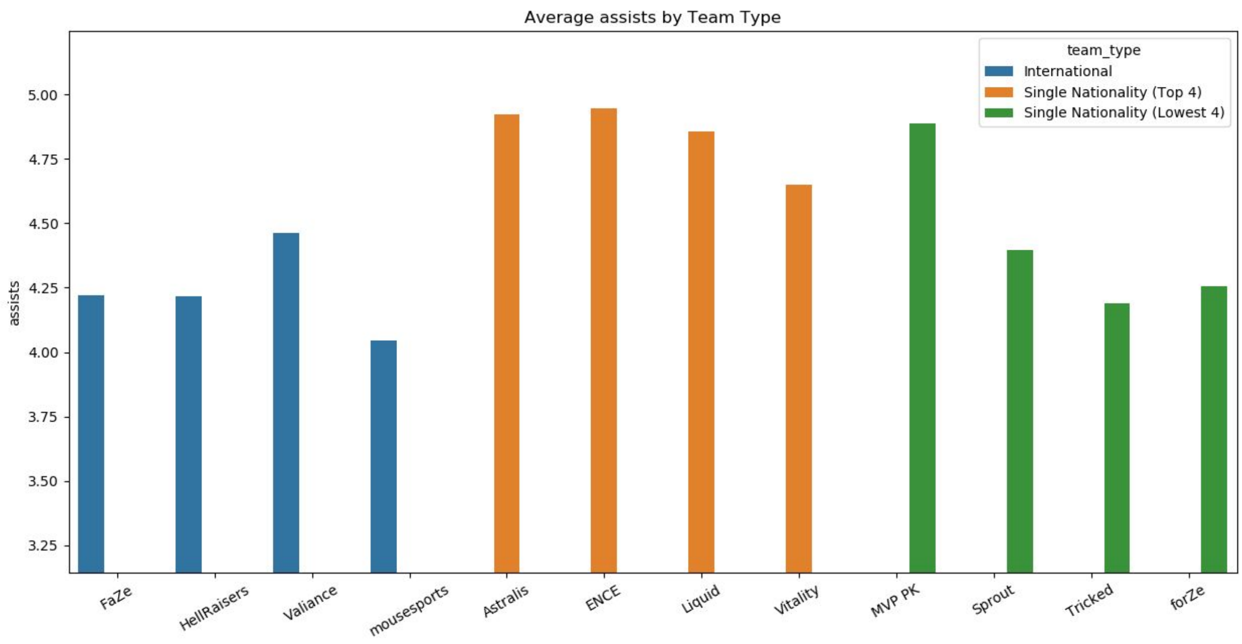
The majority of top 30 teams also seem to have greater distribution in performance when it comes to kills, opening kills, and successful flashbang grenades. This is most likely due to each team being built as a unit of differing skill sets, with a few players dedicated to shooting things to death (kills) and other players being dedicated to making decisions and using utility, like flashbangs.

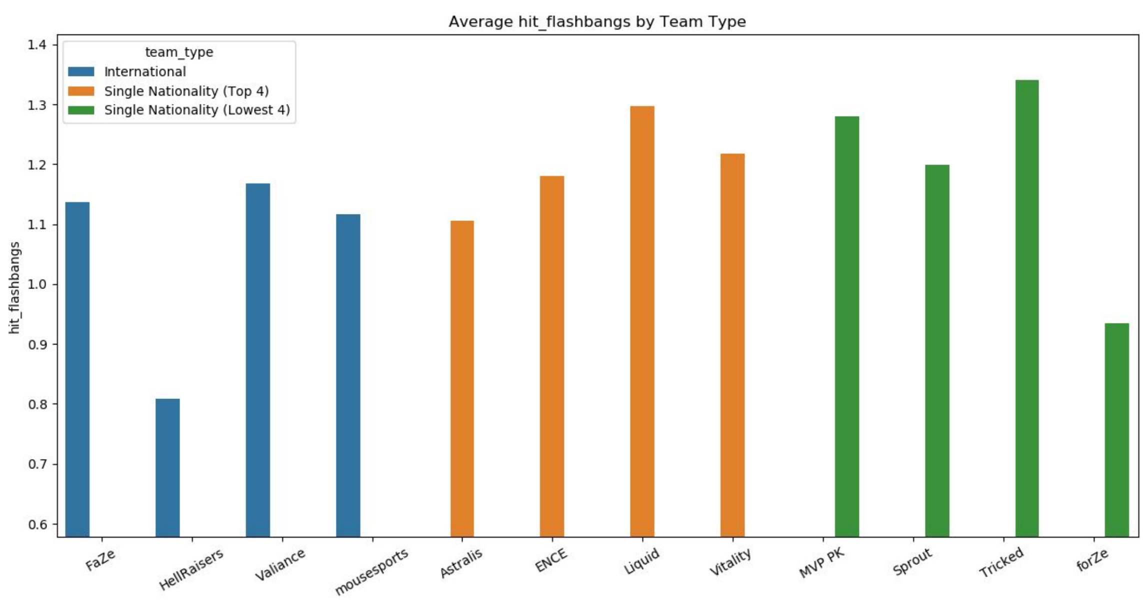


Stats like ADR and assists also weren't as equally distributed as KAST along each player on each team, but still fluctuated less than opening kills and the such.

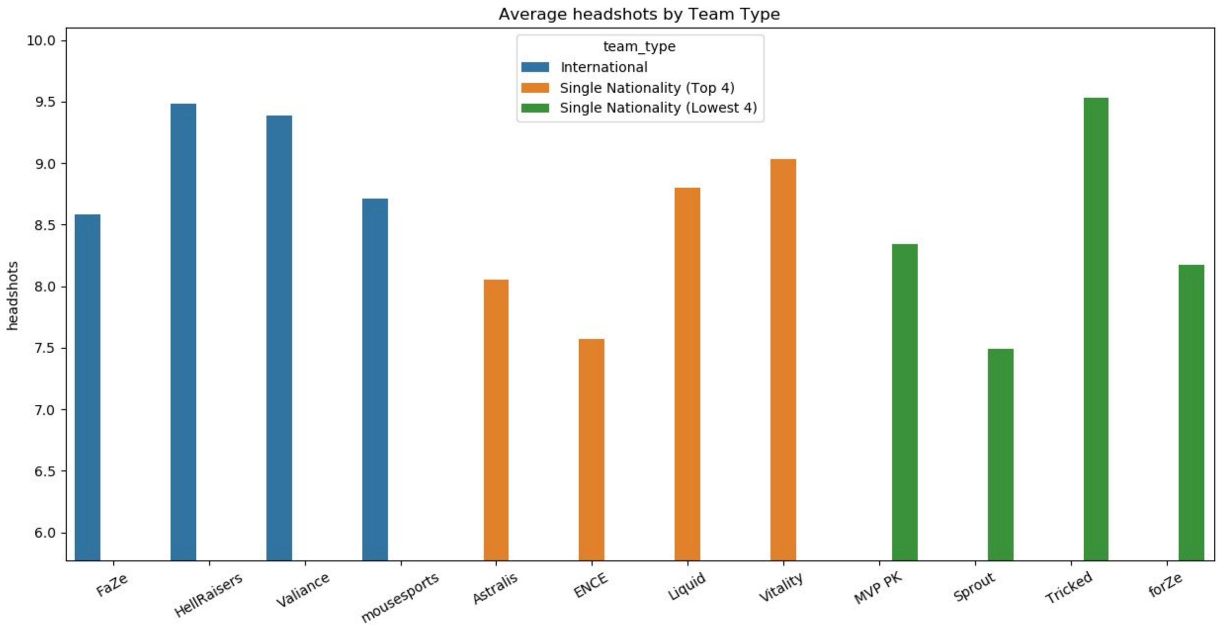


We also took a look at how international teams (Teams with no ethnic majority in player composition) compared to single nationality teams (Teams with 3+ players from the same country/region). Stats that would make sense to be lower due to language barriers, less cohesion and teamwork, such as hit flashbangs and average assists were lower. We used bar graphs seperated by team type: international, the top 4 national teams, and lower 4. Some graphs are shown below, all can be found in the “compare” folder.

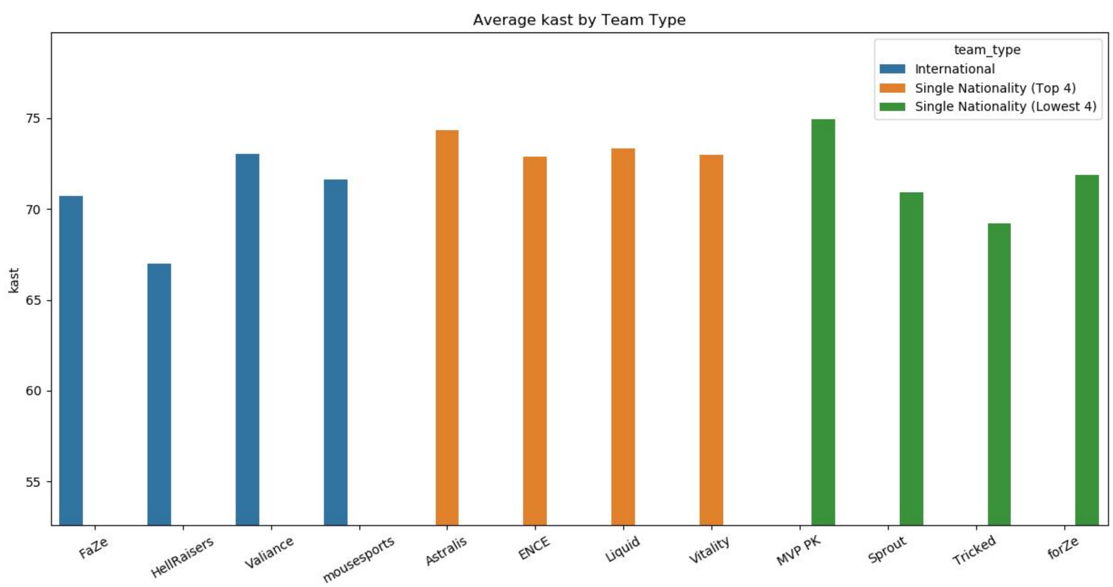




However, a lot of these international teams are built on the idea that when restricting a team's construction to one nationality, you restrict how much talent you could possibly have, whereas with international teams, you can pick the best of the best for your team from all regions. Therefore, it makes sense that international teams might score higher in terms something based on skill, like headshots.

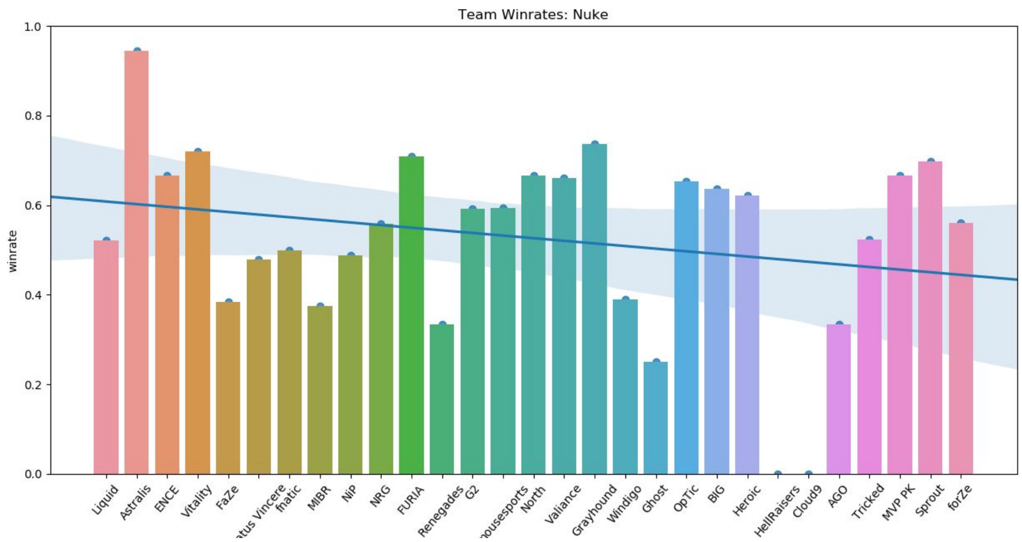


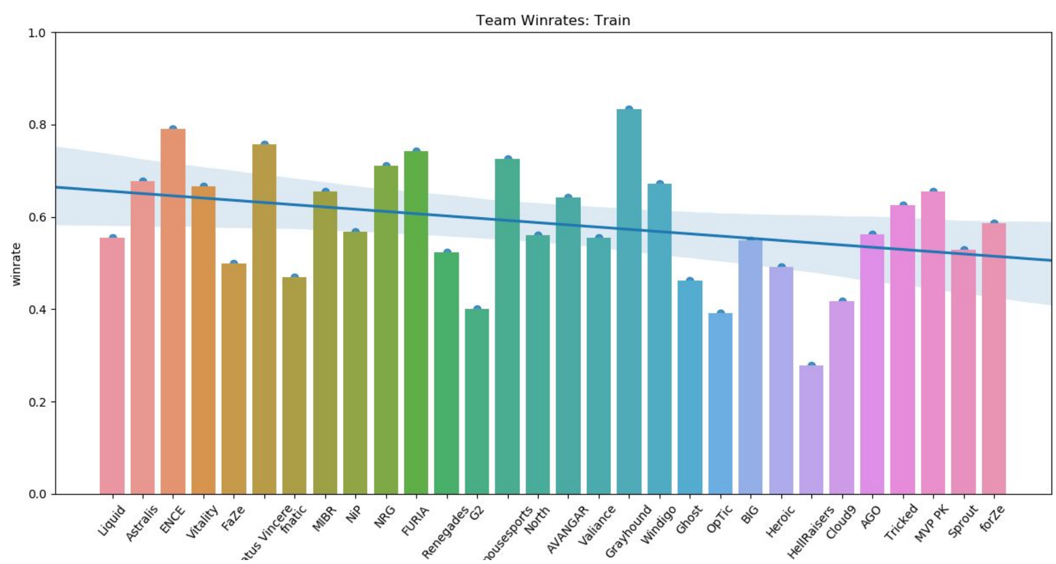
They also seemed to remain on par with single nationality teams in terms of ADR, kills, opening kills, and KAST.



When looking at win rates per map for each team, trends were a little harder to find. The way we modeled our graphs made it very obvious that regardless of which map was played, usually the higher ranked teams would have a higher win rate, which makes sense because that's how winning works. However, we could still derive a few claims from the data, which can be found in the "map_winrate" folder. For instance, higher ranked teams don't always have a higher win rate on each map. Since CSGO has a map-ban-pick system where teams rotate banning and choosing maps to play against each other, smart teams can avoid playing their opponent's historically good maps against them. Team Liquid, for example, has a significantly lower win rate on Nuke than Astralis, and more often than not ban it when playing against them. Some other trends we found were that most teams in the top 30 had higher win rates on maps like Inferno, Train, and Overpass, which implies that they are more popular than less played maps like Dust2 and Nuke.

As you can tell from the graphs, Inferno has a wider range of teams maintaining a high win rate than Nuke. Also, it is interesting to see how Astralis might be at their #2 spot due to such a staggering win rate advantage on Nuke.

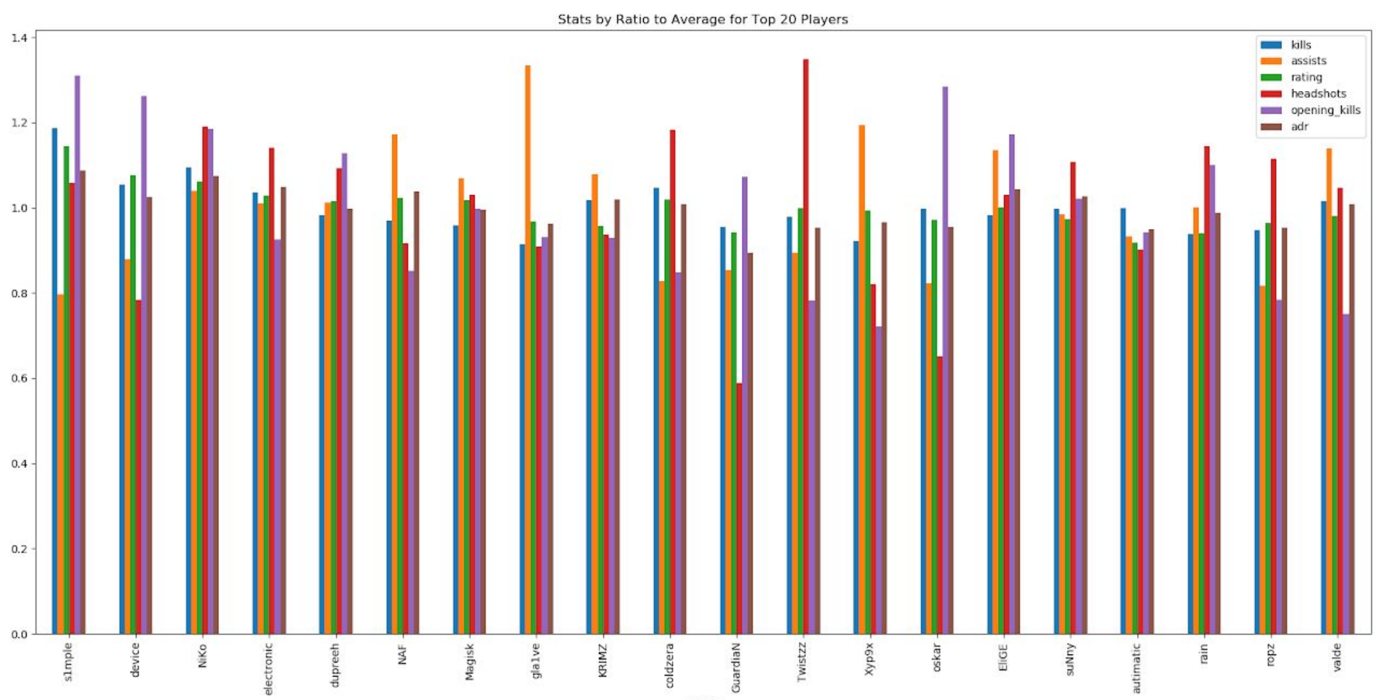




2. The second question is very similar to the first, starting with grouping player stats and finding the mean to get the average statistics per match. From there it's a matter of plotting their ranking against each of their other stats to see which is the most correlated with their rating. We can compare this to their online ranking to see how deterministic each stat is.

   Our method uses groupby with the player names to find their average statistics across all matches, then filtered to only include what most people consider the current 20 best CS:GO players using online ranking. We filtered out columns we didn't need and used loops to find the average of each stat between all players. Then, each players' stats are divided by the average to see how far ahead or behind they are of the average. The results are plotted below.

Any stat with spikes in highs and lows means it wasn't consistent among top players. - these stats are likely less important for better play because they aren't something everyone has. Kills is one of the most consistent columns, and slowly decreases as you go from left (best players) to right (still elite, but worse players). Average rating is obviously also very consistent, since rating is determined by HLTV.org based on an algorithm that uses overall performance, and these are HLTV's top 20 players of 2018. The most consistent category on the graph besides rating seems to be ADR, or average damage per round- it makes sense that top players consistent output a high amount of damage every game. Opening kills is an

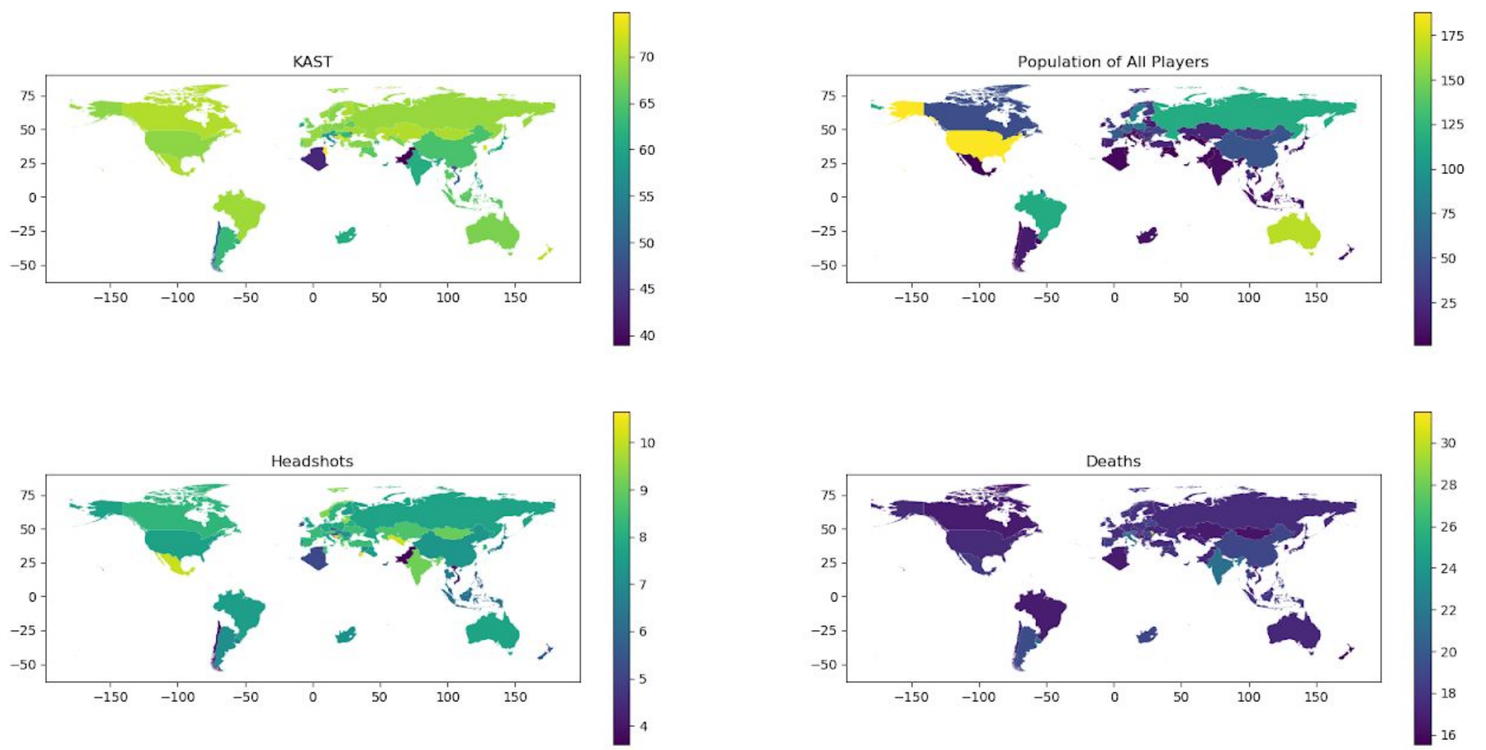


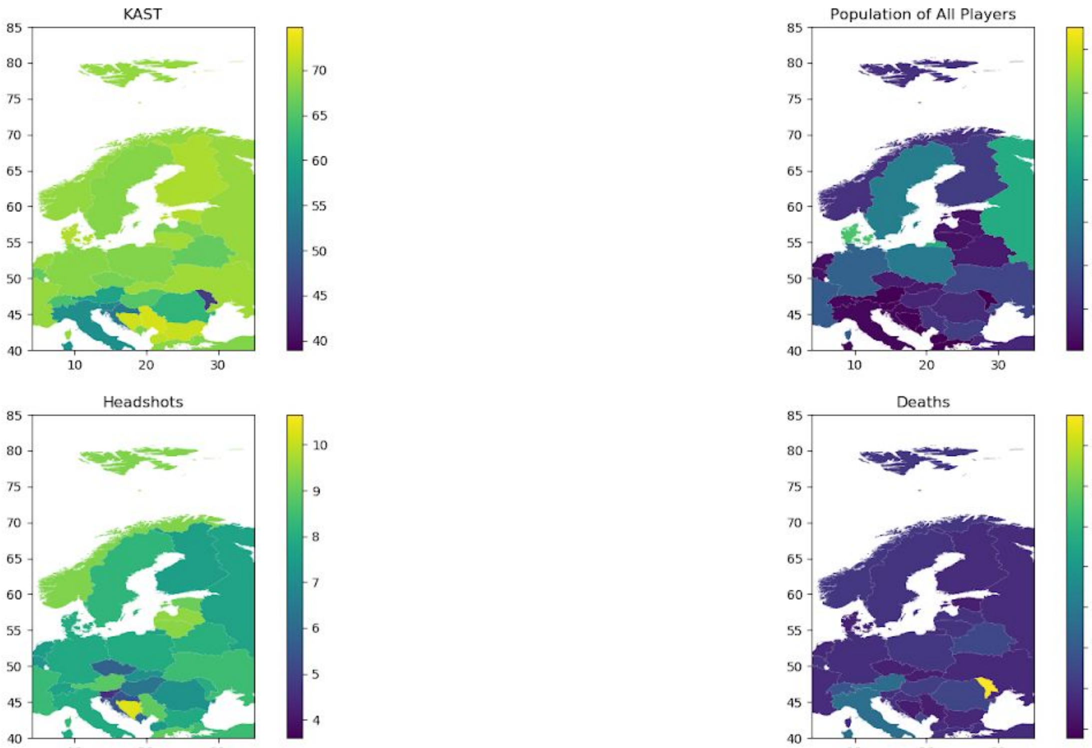
interesting category, as all top 5 players score very well in it in addition to oskar, demonstrating how a large factor in a player's success is determined by how well they can open up rounds each game by finding opening kills. This data also seems to show player archetypes, as they seem to specialize in a few stats around a certain playstyle. The most elite are all players who specialize in kills and opening kills- demonstrating that it is not always precision in combat (headshots) or teamwork (assists) that make someone better, but rather your ability to create opportunity for you team and win firefights. However, this is not to say that these aren't important factors. Twistzz, of Team Liquid, is a good example of a standout player in terms of headshots, even among the top 20 players of 2018, he soars in terms of precision. Gla1ve in another example, him holding the in-game-leader (team captain) role of Astralis puts him in a position to make smart decisions for his team, rather than to search out kills. This is illustrated nicely in his above average assist rating compared to the other top 20.

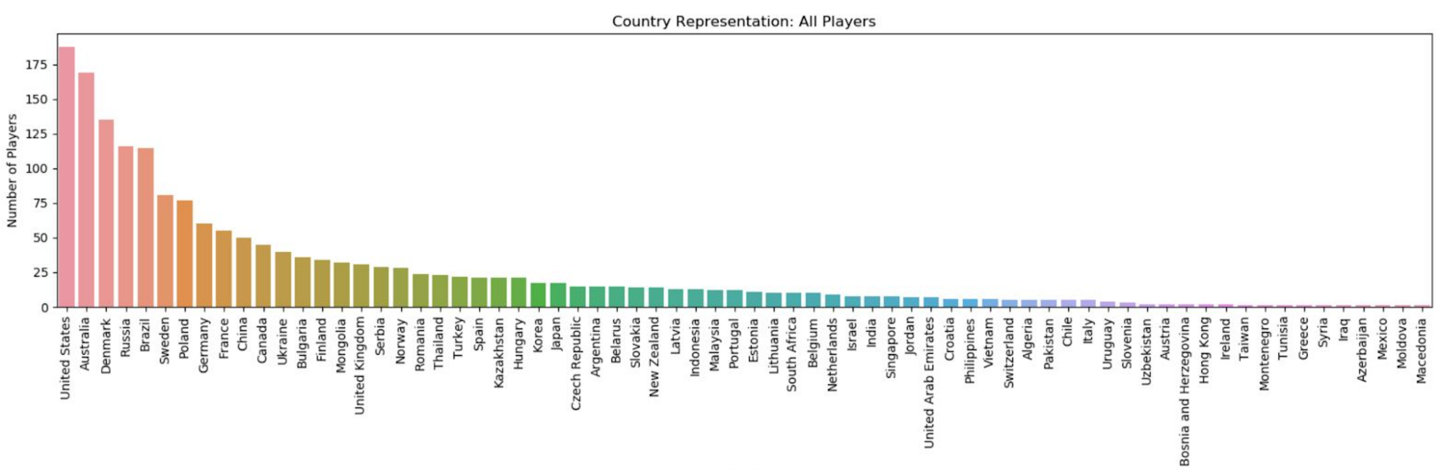
3. In looking at what countries produce the strongest players, we can first look at our player data after each stat is averaged for each player. From there, we use their country of origin to group them again and again find mean statistics. We can see what countries average the highest stats. We can also take the same data and count how many unique players are from each country by looking at the grouped data. A bar chart next to a different bar chart of where the tournaments are located/country they are in can show what countries are the most heavily invested in the game and produce the most pro players. Additionally, we can examine which regions have the most representation at large Valve-sponsored tournaments. These tournaments are considered the most important per year, and by grouping by these tournaments, we can see which teams step up at these important times.

   In our program, we create a duplicate of our player dataframe, take out any duplicate names and count each player for each country. We merge this with the country geometry dataset, then drop any countries we don't have data for. The average stats per country is calculated with the original player/match dataset, finding the average stats per match of each country. This is combined with the already combined dataset, and two images are produced. Each has four graphs, each highlighting a particular stat (KAST, population of players, headshots, and deaths) but the second image is changed to only display data of scandinavian countries, which are small and hard to see in the original plot but filled with info. The plots are on the following pages:

Data with all Players: WorldPlayerStats.png , ScandinavianPlayerStats.png, country_rep/country_rep.png

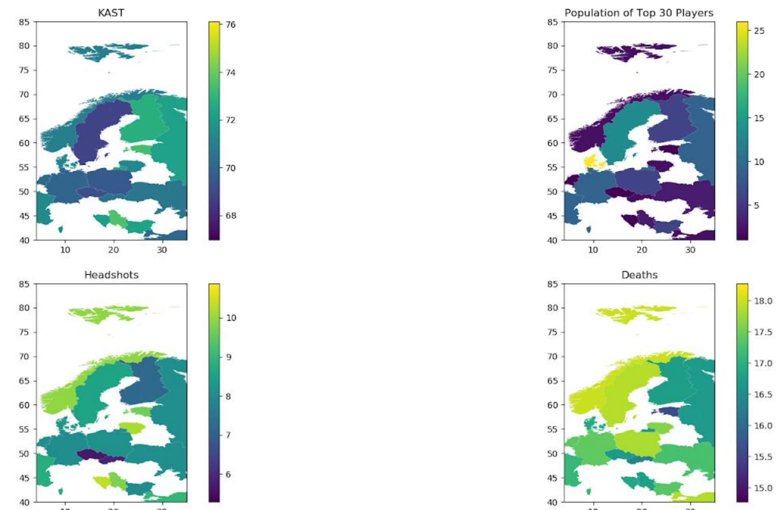


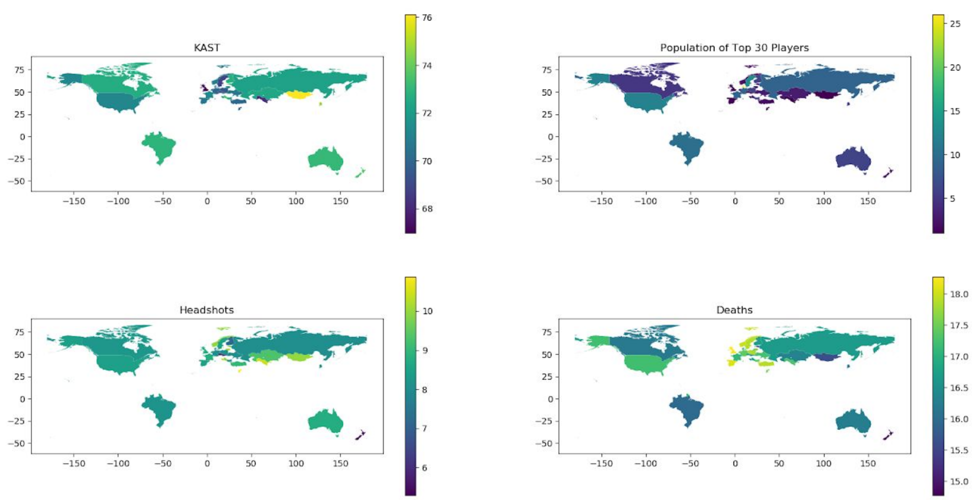

KAST
Population of All Players
Headshots
Deaths

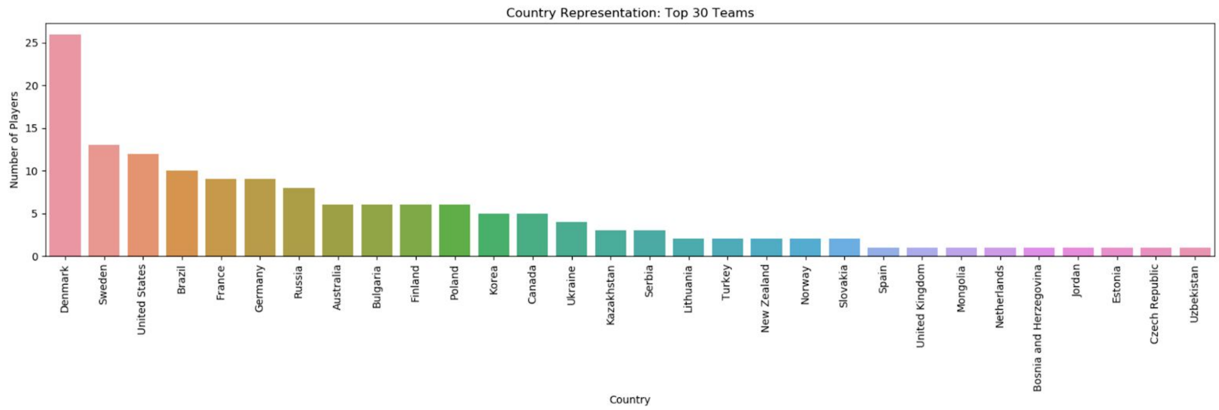

Country Representation: All Players
Number of Players
Country
United States
Australia
Denmark
Russia
Brazil
Sweden
Poland
Germany
France
China
Canada
Ukraine
Bulgaria
Finland
Mongolia
United Kingdom
Serbia
Norway
Romania
Thailand
Turkey
Spain
Kazakhstan
Hungary
Korea
Japan
Czech Republic
Argentina
Belarus
Slovakia
New Zealand
Latvia
Indonesia
Malaysia
Portugal
Estonia
Lithuania
South Africa
Belgium
Netherlands
Israel
India
Singapore
Jordan
United Arab Emirates
Croatia
Philippines
Vietnam
Switzerland
Algeria
Pakistan
Chile
Italy
Uruguay
Slovenia
Uzbekistan
Austria
Bosnia and Herzegovina
Hong Kong
Ireland
Taiwan
Montenegro
Tunisia
Greece
Syria
Iraq
Azerbaijan
Mexico
Moldova
Macedonia

KAST (Percentage of rounds where a player either gets a kill, an assist, survives the round, or is responsible for another enemy killed via death by distraction) pretty consistent everywhere, with Europe and some of NA having an edge. Some outliers appear from countries that have one or two elite players only. Population is dominated by the US and Australia on a world level, but interestingly enough Denmark is one of the most concentrated player bases in the world despite its smaller size and civilian population. Looking at headshots, the results there aren't too unique results, Norway and Mexico seem to be fairly strong in that department. China, known for their accuracy, doesn't have the most headshots in our result. This is likely due to the fact that China having less professional teams who bring the average down. Deaths seems to be fairly consistent, with a few outliers. One observation we can make from multiple plots is about imported talent. The outliers that appear in the plots come from a few individuals coming from less populated countries. This is likely due to the fact that these individuals came from places with less CS:GO players, but were good enough at the game that they moved and joined a professional team. In this way, there is more of a filter for talent for people coming from these countries with less players. It also good to note that for these above 2 figures, we are using our entire dataset, which is scraped in a way that doesn't include the same number of rows for certain players from lower tier teams as for higher ranked ones. To gain a better understanding of how country representation is dispersed among players, we can filter our data by only including players on the top 30 teams.

Data with only Top 30 Teams: World30Stats.png , Scandinavian30Stats.png, country_rep/country_rep.png

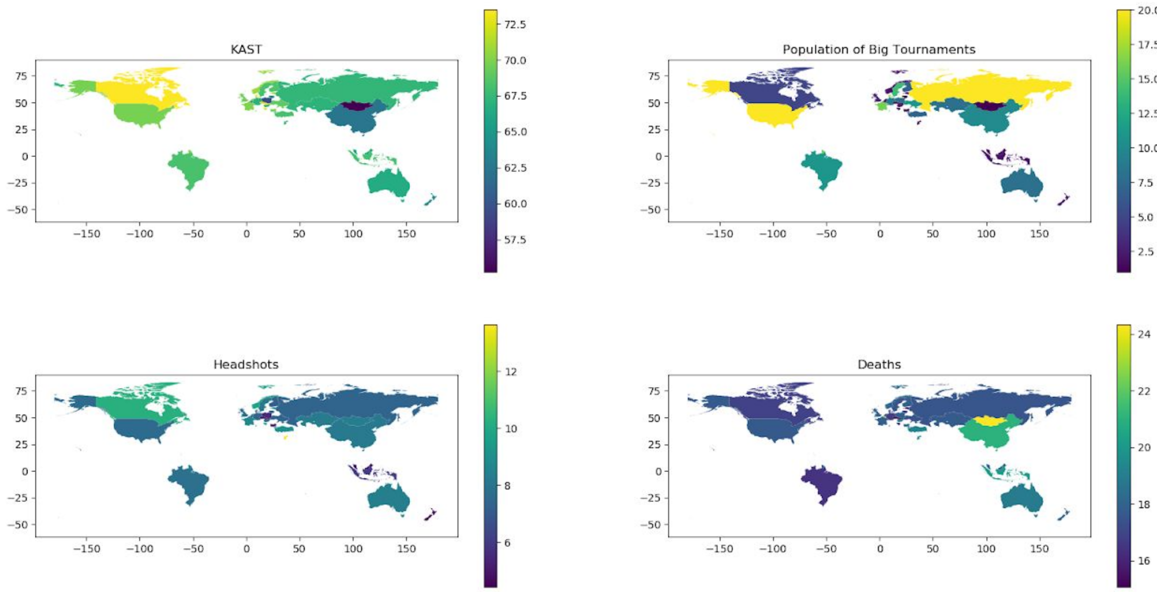


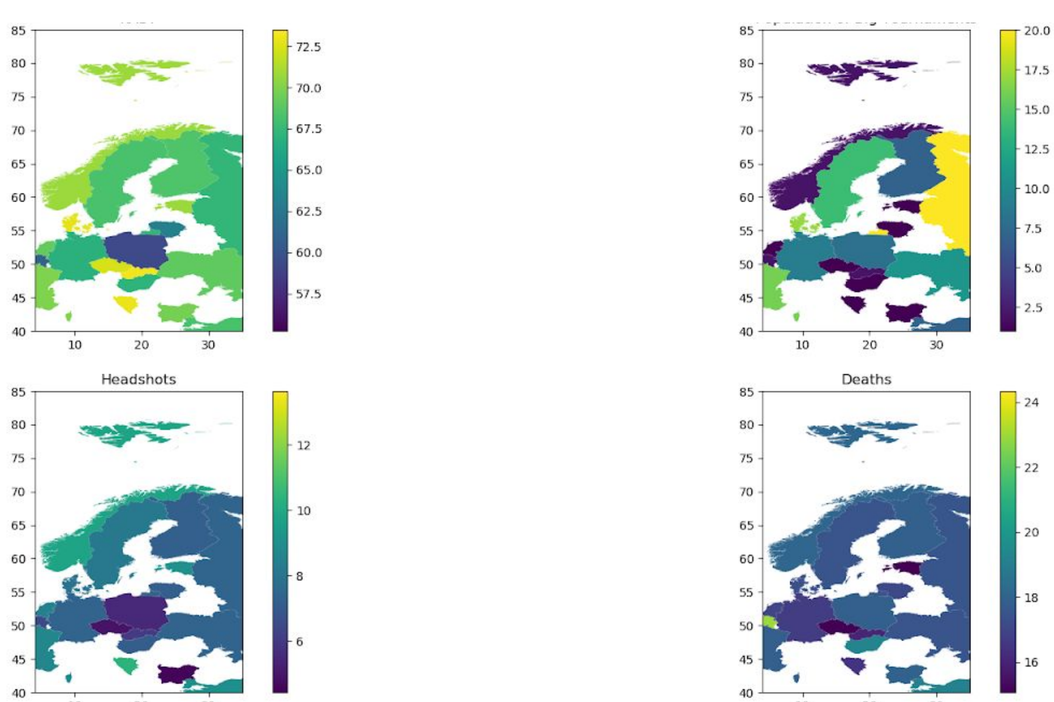


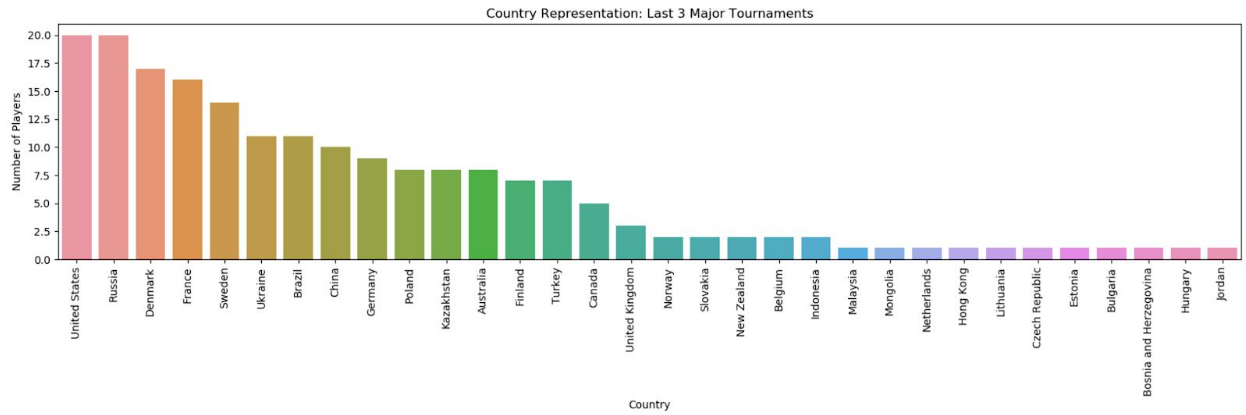


Here's the same information, but only looking at players on the starting lineup (5 people) of a top 30 team. Finland and NA seem strong in the KAST category, population is dominated by Denmark, Norway is good in headshots, and deaths is usually led by Western European countries. It's interesting to note how in a smaller player pool, countries like Bosnia Herzegovina and Jordan really stand out with only 1 player representing their population. Denmark has the most talented players, followed by Sweden and the US. This data is a lot more dynamic here because it's only looking at 150 players. It's overall less representative, but it's important to look at the big time players, especially in population, to get a feel for the strength of each country.

Data from Last 3 Majors: WorldBigTourneyStats.png , ScandinavianTourneys.png, country_rep/country_rep.png



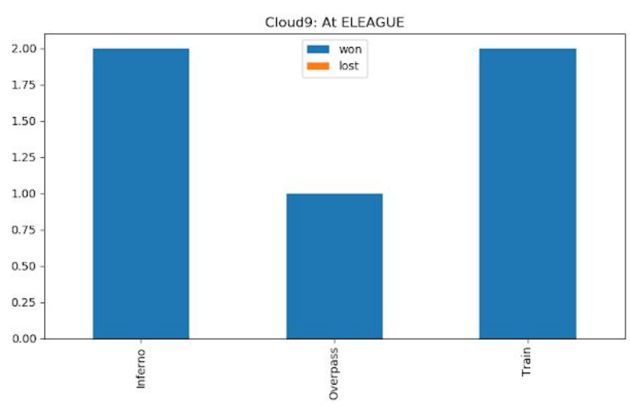


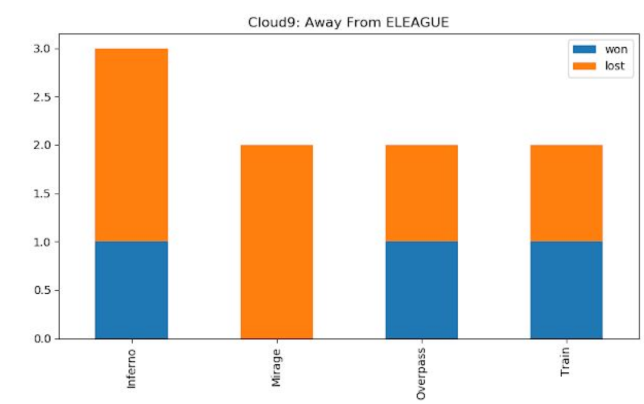


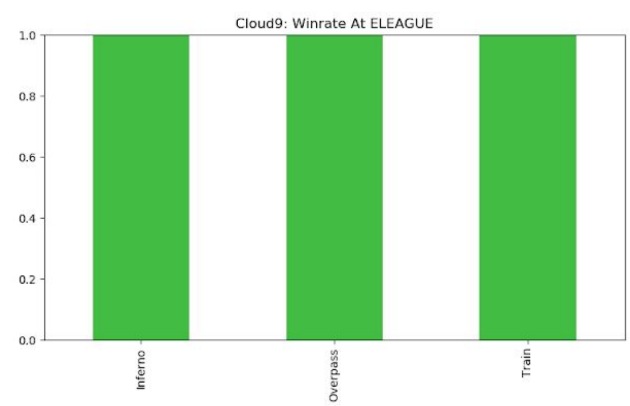
Finally, we plotted all the same info but only using data from the past three *Major* tournaments. These tournaments are the biggest of each year, and should be representative of which nationalities reach peak performance at the correct times. The US and Russia dominate in terms of unique players at these tournaments, followed closely by Scandinavia and France. In terms of stats, a lot of the same countries seem to be performing relatively the same even when looking at these parts of the group, especially Denmark. Sweden and Denmark also have a surprisingly large player base at these larger events. Norway and the other Scandinavian countries are still ahead in headshots. Asian players seem to show up at these tournaments more often than they show up in the top 30 list, and have way more deaths at these large tournaments per match which might be due to a flawed qualifier system that lets in a quota of teams per region.

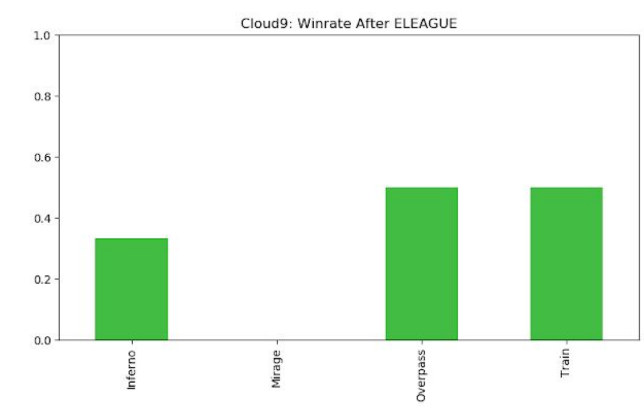
4. For home field advantage, we can use our tournament dataset to get tournament locations, and examine player performance in venues located in their home countries versus not. Specifically, we could look at Cloud9's, a North American Team, performance during their successful Major run in 2018 at Boston and compare that to their matches leading up to the tournament and after the tournament. Was their success a fluke? Similarly, we could look at Made in Brazil's performance at Sao Paulo, or Renegades performance at IEM Sydney 2018 and 19. By separating by "at home" and "away from home", and then aggregating team stats, it should be simple to visualize the effect of home field advantage.

We came to this question assuming that more often than not, teams would perform better with a home crowd cheering them on. This was from my past experience watching North America's Cloud9 win the ELEAGUE Major tournament in 2018. However, that seemed to be just a fluke. Cloud9 Looked significantly better at ELEAGUE than they did at the next few tournaments after their victory. These graphs can all be found in the "tournament_winrates" folder.

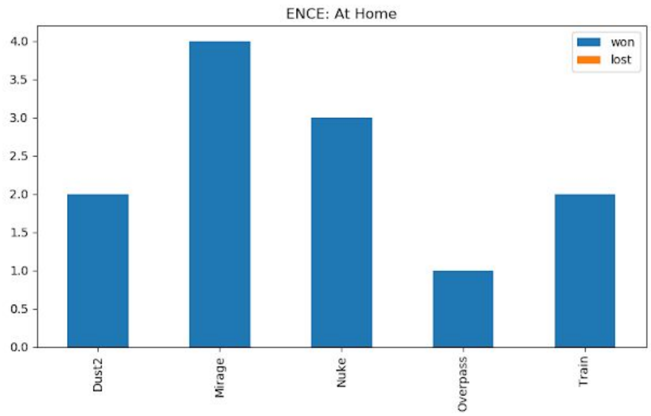


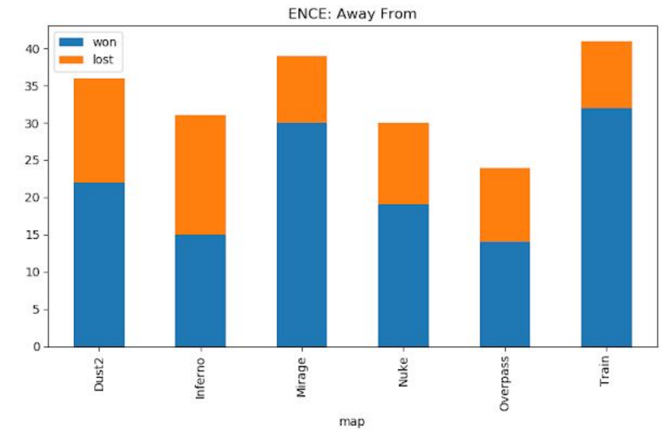


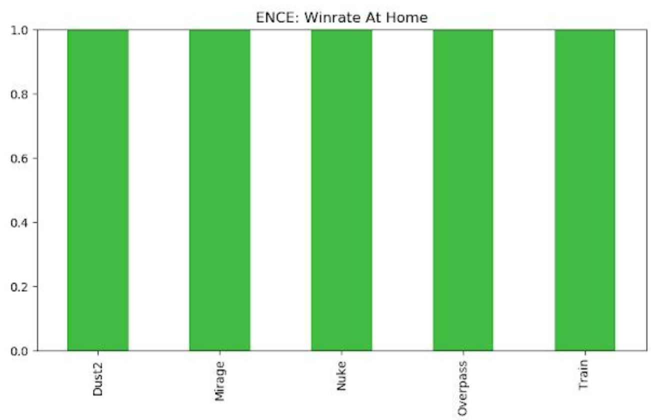


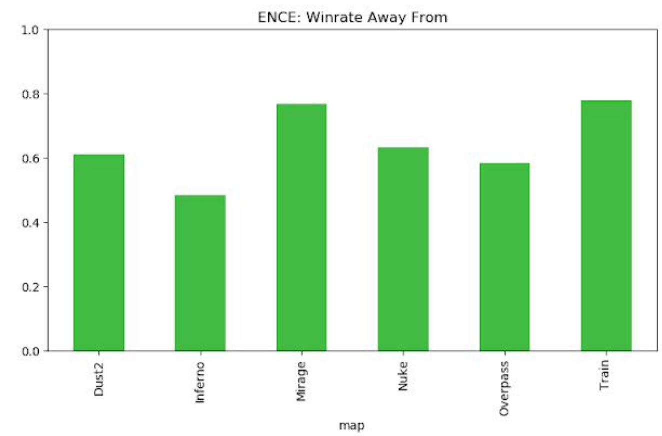


They maintained a clean record during the tournament but fell apart afterwards at tournaments that took place in Poland, Haikou, and Kiev. Similarly, teams like ENCE, from Finland, and TYLOO, from China, maintain a consistently higher win rate at tournaments in their respective countries than outside.



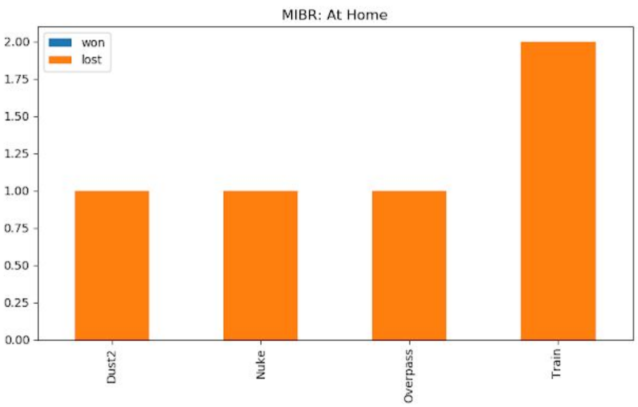


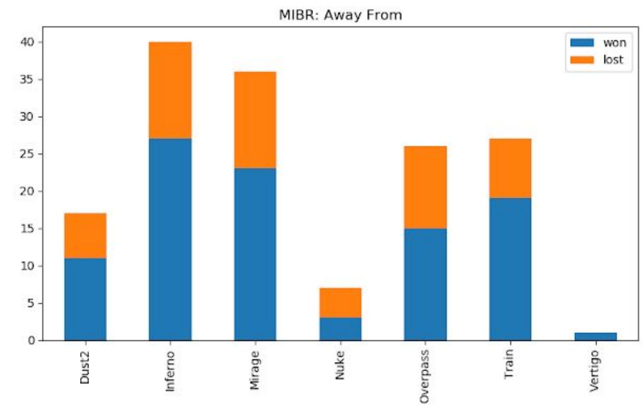


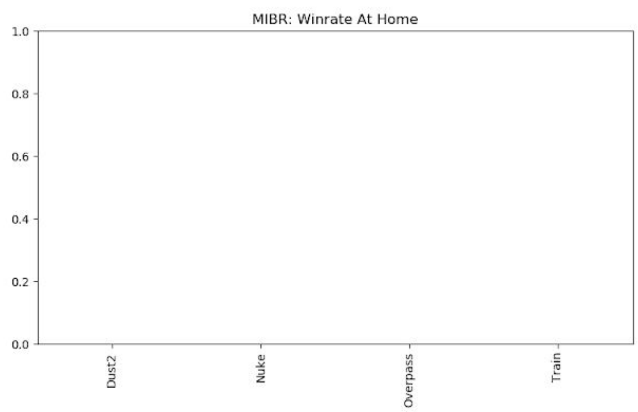


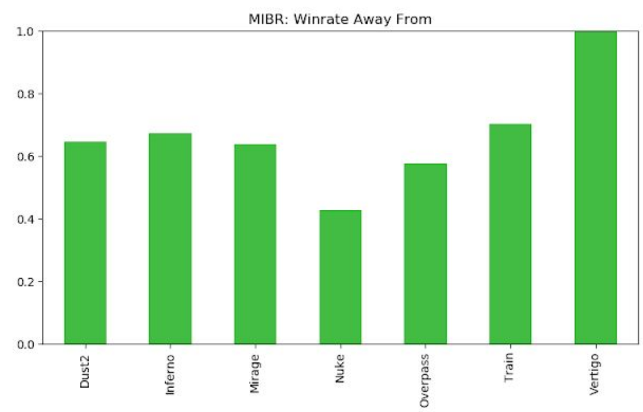
However, due to the findings of the next graph and with prior knowledge of TYLOO and ENCE, these results are most likely due to each of the teams being the strongest in their regions, and thus having an advantage over other teams that play in smaller local tournaments as opposed to global ones.

We also took a look at MIBR's (Made in BRAZIL) performance in Sao Paulo this year and found some surprising results:

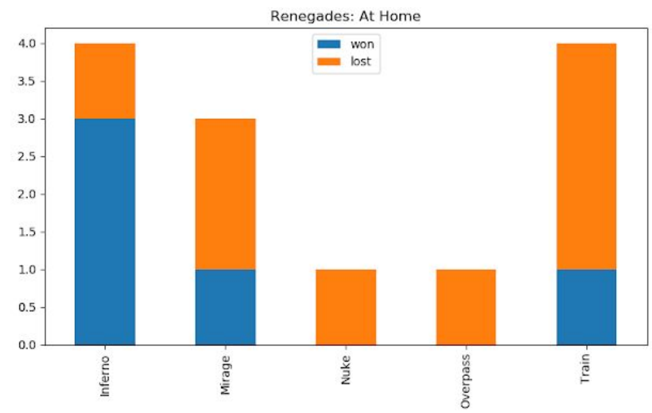


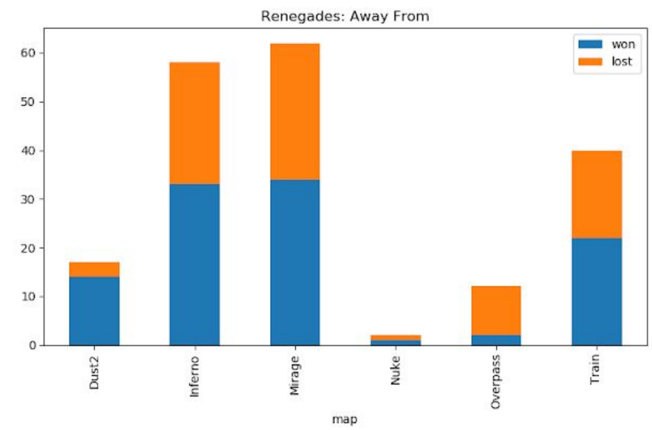


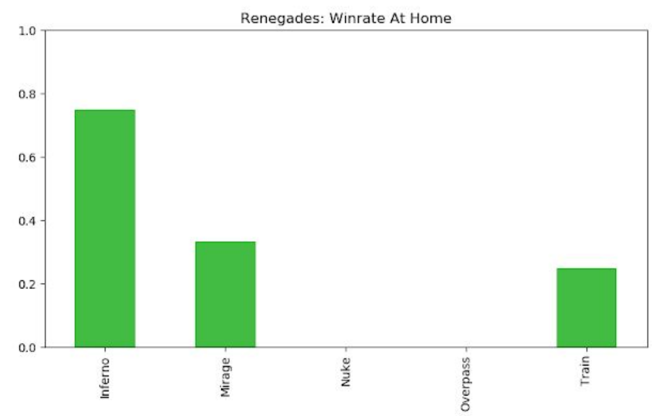


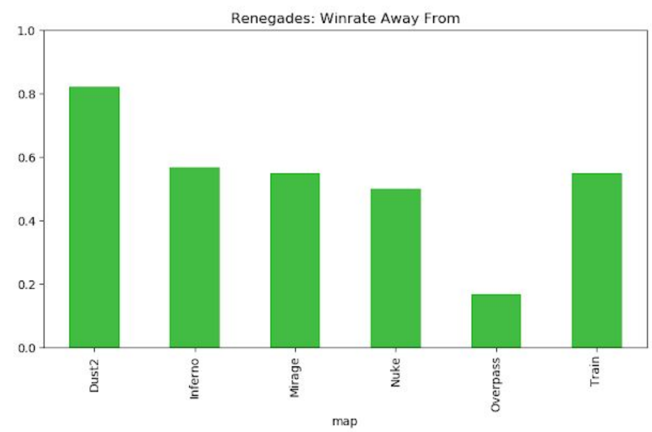


MIBR didn't win a single game in Sao Paulo out of the five they played, but have always been considered one of the best teams in the world. It seems their home crowd made them play worse, if anything. Similarly, Renegades, from Australia, didn't seem to step up during their home tournaments, IEM Sydney 2018 and 2019. They maintain a higher win rate on maps like Mirage and Train outside Australia.









These results are solid enough to make the claim that teams don't usually experience a home field advantage, and that Cloud9's sudden victory in Boston was most likely due to luck and hard work. MIBR and Renegades prove that even though crowds may be loud and supportive (the Australian crowd in Sydney was rowdier than anywhere else), teams still might not perform better than usual. This data is also not perfect. We of course don't take into consideration the level of teams that these teams are playing against. At Sao Paulo, MIBR was taking on the top 3 teams in the world, and ENCE and TYLOO usually played 2nd tier teams when in their home country.

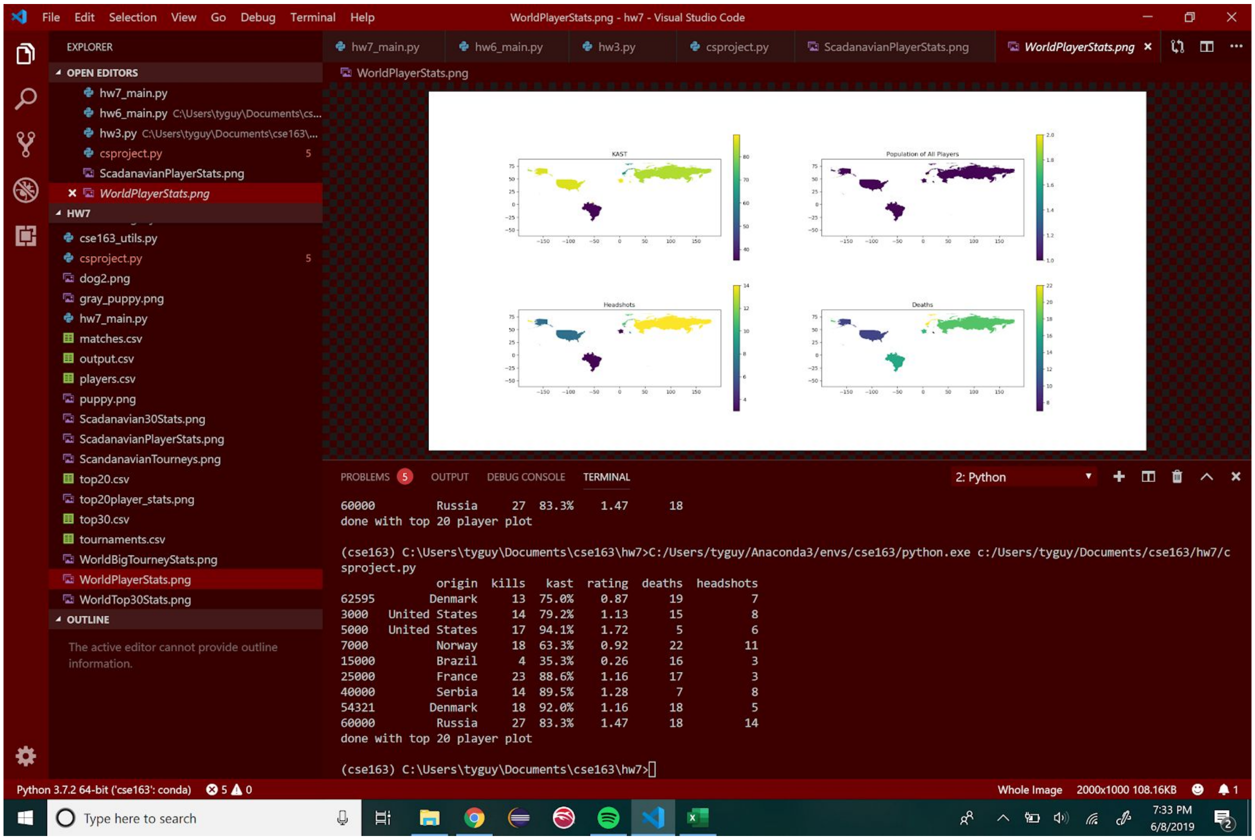
Reproducing Results:

1. To get our program and requisite folder, you can download our repo at https://github.com/ryansiu17/csgo_py
2. You will need the libraries pandas, matplotlib, seaborn, numpy, geopandas, and pycountry.
3. You will also need the datasets:
    a. matches.csv
    b. players.csv
    c. top20.csv
    d. top30.csv
    e. tournaments.csv
    f. countries.geojson
4. From there, running questions_1_4.py and questions_2_3.py should be enough to generate graphs of all our data.
5. If there is a missing file error, make sure you have the folders
    a. advantage_winrate
    b. average_per_match
    c. compare
    d. country_rep
    e. map_winrate
    f. stat_dist
    g. stat_per_team
    h. tournament_winrates

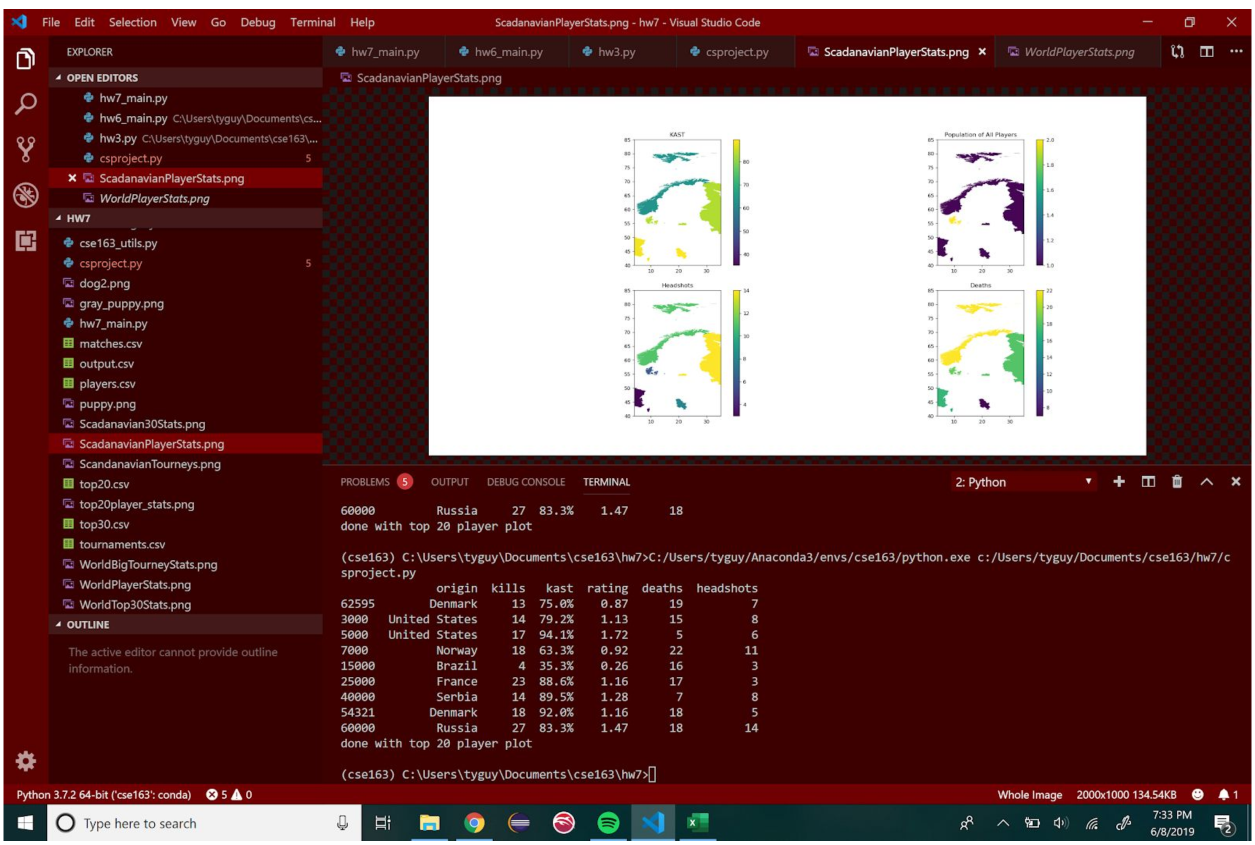
Evaluation of Work Plan:
Fortunately, out data was collected and mostly ready during part 1. Having to scrape a few more categories took a small bit of time, pushing our experimenting period a day late. Experimenting with our data the first few days was fun and easy but writing finalized code took a while after. Test functions were written later than previously planned, as was the report. Writing good, working code was harder and took significantly longer than expected, costing a few late days. Reporting the data was easier and allowed us to get back on schedule.

Testing:
To test our questions_2_3.py data, we used a small piece of the original data where we can manually find the results quite easily. To test our tests, you can uncomment the lines near the to of the file that have descriptions of what tests they are. As you can see from the screenshots, the information from the handful of players we looked at was mapped properly to our plots. The French and Serbian player we nearly tied

for highest KAST, and thus both have a yellow color, showing they are at the top end of the spectrum. Each player in the test data was from a unique country, except Denmark, and the graph displays this too. The Russian player had the most headshots and Brazil/France had the least, as reflected by the graph. Norway leads in deaths. We know this takes the average because Denmark's KAST in this data is roughly low 80s, right in the middle of the 75 and 93 that the dataset contains.

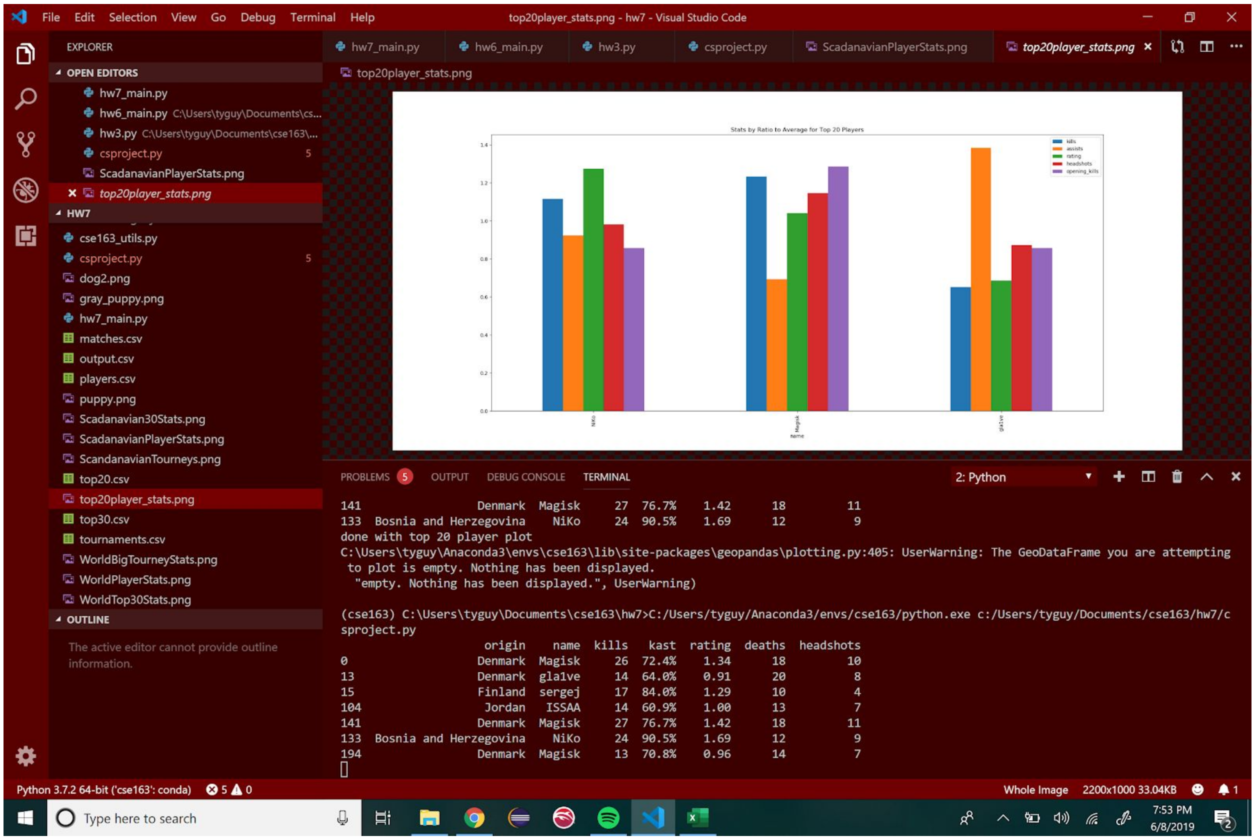


WorldPlayerStats.png - hw7 - Visual Studio Code
KAST
Population of All Players
Headshots
Deaths
PROBLEMS 5 OUTPUT DEBUG CONSOLE TERMINAL
60000 Russia 27 83.3% 1.47 18
done with top 20 player plot
(cse163) C:\Users\tyguy\Documents\cse163\hw7>C:/Users/tyguy/Anaconda3/envs/cse163/python.exe c:/Users/tyguy/Documents/cse163/hw7/csproject.py
origin kills kast rating deaths headshots
62595 Denmark 13 75.0% 0.87 19 7
3000 United States 14 79.2% 1.13 15 8
5000 United States 17 94.1% 1.72 5 6
7000 Norway 18 63.3% 0.92 22 11
15000 Brazil 4 35.3% 0.26 16 3
25000 France 23 88.6% 1.16 17 3
40000 Serbia 14 89.5% 1.28 7 8
54321 Denmark 18 92.0% 1.16 18 5
60000 Russia 27 83.3% 1.47 18 14
done with top 20 player plot
(cse163) C:\Users\tyguy\Documents\cse163\hw7>



ScadanavianPlayerStats.png - hw7 - Visual Studio Code
KAST
Population of All Players
Headshots
Deaths
PROBLEMS 5 OUTPUT DEBUG CONSOLE TERMINAL
60000 Russia 27 83.3% 1.47 18
done with top 20 player plot
(cse163) C:\Users\tyguy\Documents\cse163\hw7>C:/Users/tyguy/Anaconda3/envs/cse163/python.exe c:/Users/tyguy/Documents/cse163/hw7/csproject.py
origin kills kast rating deaths headshots
62595 Denmark 13 75.0% 0.87 19 7
3000 United States 14 79.2% 1.13 15 8
5000 United States 17 94.1% 1.72 5 6
7000 Norway 18 63.3% 0.92 22 11
15000 Brazil 4 35.3% 0.26 16 3
25000 France 23 88.6% 1.16 17 3
40000 Serbia 14 89.5% 1.28 7 8
54321 Denmark 18 92.0% 1.16 18 5
60000 Russia 27 83.3% 1.47 18 14
done with top 20 player plot
(cse163) C:\Users\tyguy\Documents\cse163\hw7>

For looking just at top 20 players, we ran data filtered to 5 rows, 3 containing info on the same player. The result is as follows:



As you can see in the console, we had some players not in the top 20 in our test dataset. They were filtered out as expected. NiKo had the best rating both on the graph and in the test dataset. While Magisk had more kills than NiKo in two out of his three performance, his last match with only 13 brings his average under NiKo, which is the result we see on the graph. Overall, the results of both graphs are exactly what we would expect from the test data. Because we also made sure it would average in both cases, it is fair to assume this program works properly on a much larger version of the test data, which is what we give it. Given our program, these results can be recreated by uncommenting certain lines that filter the larger dataset into the test data.

To test questions_1_4.py, we used another smaller set of data, testing the more in depth functions that weren't as easy to check simply by going through code on paper. There is a file, test_1_4.py that contains the tests for those few functions. Most testing was done on an online colab where prints were used during experimenting.

Presentation
Congrats! You get to listen to both of us talk for a collective 4 minutes on Tuesday about our project! (We will present in class)

Collaboration
AUTHOR2 came up with the idea and scraped the dataset together. He suggested some research questions and added extra sub-questions and formatting to our plan. He handled the successful player and home field advantage questions as well as the bar graph for country representation.

AUTHOR1 came up with a lot of the research questions and sub-questions with it. He handled the investigation of key stats in top 20 players and which country produced the strongest players, and in what category. He also wrote this.