

Compiling and Presenting Ultimate Data

Timothy Newman and Benjamin West

May 2016

1 Abstract

In this paper we present a tool to visualize the data publicly provided by one of Ultimate Frisbee's professional leagues, the American Ultimate Disc League (AUDL). To our knowledge, there have only been a few attempts to visualize this data and there are zero efforts to produce some of the most basic visuals. We have compiled all the data in one place and wrapped it in a nice tool to allow easy exploration.

2 Introduction

Having detailed statistics for Ultimate is a very new concept[1]. USAU, the governing body for club ultimate, has tracked some basic statistics like goals and D's for a little while. These are limited, and they generally only track them during Nationals, which leads to a small sample size.

In 2012 two club leagues started - the American Ultimate Disc League (AUDL) and Major League Ultimate (MLU). With these has come far more detailed records of games. So far the only people we have seen explore

this data have been the leagues themselves. For example, the AUDL displays the individual statistics for each player, a large and cluttered parallel coordinates system of all the players, and a visual of how often players passed to different team mates.

What is lacking is a way to really explore these statistics in the basic ways we expect to be able to. Other sports have a wide variety of ways to do this, from compendiums of every game ever recorded[2], to news sites dedicated solely to covering them[3]. We aim to build one of the first tools in Ultimate to do this. Hopefully, this will allow players to ask and answer basic questions such as "Who is the best at scoring?", "What are the standards for good players?", etc, and challenge the common wisdom in the sport today.

To this end we have compiled all the data from the AUDL and wrapped it in a nice front end that allows for easy generation of basic graphics such as scatter plots and bar graphs. Even though we only have the basics done it has already prompted us to ask and answer questions we had never considered before even though we have a collective 6 years of experience. Once our tool is a little more

polished we plan to release it to the general community and garner feedback.

3 What Is Ultimate?

Before we get too far it's important to give some context. Ultimate is a field sport played with seven players per side. Just like in football, you score by catching the disc in the endzone you are attacking. Players may not run with the disc, and it is advanced by passing it to your teammates. This must be done within ten seconds, or it is a turnover. If the disc touches the ground while a player is not holding it, only the other team may pick it up. To start a point, one team throws to the other - like a kickoff in football.

These rules result in statistics frequently seen in other sports such as basketball, football, and soccer. The current big four statistics that get the most talk are goals, assists, Ds (when you cause a turnover on defense), and turnovers. There are stats that everyone knows about but less people track because it is more time consuming such as number of throws and completion percentage. Lastly, Ultimate has some plays unique to itself. An example is a Callahan, which is when you score off of the other team's throw.

The bottom line is that while Ultimate has a unique combination of stats, most of them are pretty similar to other well-known general sports metrics. There are not yet any statistics that are of comparable complexity to WAR from basketball or VORP from baseball[7][8].



Figure 1: Photo credit Alain Hwang. Play credit Andrew Yale, Data Structures TA

4 Formulation

Before we describe the tool, it is important to describe our exact motivation for creating the tool. Understanding why we created it and who we created it for should explain some of the decisions we made.

4.1 Motivation

We touched on this briefly above, but it is worth elaborating on. Ultimate is in the middle of one of its formative stages. Between the flood of youth programs, the creation of pro leagues, and USAU signing a broadcast deal with ESPN, it is clear that the sport is starting to gain traction in the general public. As this happens more and more people are going to want to talk about it.

We aren't saying that easy access to stats

is necessary to conversation. Before statistics in sports became mainstream, people certainly had no trouble talking about them for hours on end. But to argue that an interest in sporting numbers is only for a niche group of nerds would be completely ignoring the behemoth that is fantasy football. Knowing how the best in the sport are doing is clearly enticing for a lot of people.

We don't expect the only interest will come from diehard fans though. Ultimate is in a funny place right now where the highest level of competition is Club. This means anyone can dream of competing at the highest level, or at the very least hope to play against the best a couple times. The Ultimate community is full of extremely dedicated amateurs probably looking to improve their own game. And one of the best ways to improve is to see what the best are doing.

To wrap up why we wanted to create this tool: We expect that if there already existed an accessible presentation of currently available data that was clean and intuitive, there would be sizable audiences coming from multiple places within the Ultimate community. This is supported both by Ultimate's structure as a sport, and from experience with other sports.

4.2 Audience

Even though it may be obvious from the motivation section, it is important to define who the audience is. We are restricting to people already entrenched in the community with a good understanding of the game. We aren't intending for this to draw new players in.

Furthermore, we are restricting to people who already have some understanding of numbers and charts. It is more important to allow access to these numbers to those who might have good ideas on how to use it than it is to try and rope everyone in at once and change the mindset around the whole game. We'll get into this in the next section, but the statistics available definitely aren't the be-all end-all and there is a lot of work to be done before they are useful to a wide audience.

4.3 Goals

Our goals will go hand in hand with our motivation and our audience. We want to couple the available data to some clean and basic visualizations that are easy to create for those who already have a little background knowledge. To what specifics does this end up translating, though? We have a list below.

4.3.1 Get The Basics Done

Basic chart types should be supported. For now these include bar charts, scatter plots, and a view of all statistics on an individual player (this last one coming by our presentation on Tuesday).

4.3.2 Ease Of Creation

These charts should be easy for a user to create and alter. Without this feature, no actual exploration will occur.

4.3.3 Give Them Everything

All the data that we have parsed should be accessible within the page. We aren't trying to show people where to look.

4.3.4 Leave It Up To Them

There should be flexibility in which statistics to plot. The user should be able to decide which metrics they think are most interesting.

4.3.5 Keep It Simple

The interface should be as minimal as possible without causing inconvenience. This will not be the tool to create elaborate plots or generate complicated mathematics, and should be as easy to learn as possible.

5 Previous Work

The previous work we have seen falls into two major categories: work done on Ultimate, and work done on other sports.

Outside of the AUDL's own website there has been two projects to gather and visualize Ultimate data that we can find[4]. UltiApps is an app that allows players to track passes on the field to later get a good idea of where the disc tends to go in a typical point. It resulted in a paper at SLOAN [?] and an affirmation of some of the most basic knowledge in the game—having the disc in the center of the field is valuable. Compared to what we are trying to do, it is aimed at an entirely different area. It is for players to track their

own games, and it's best visual is a heatmap of the field.

There is also Ultistats. This is a mobile site meant to assist teams in tracking their own statistics, and produces different visuals than we do. Again, its main goal is to assist teams in keeping a record. This is a different goal from ours, which is to allow players to explore a large amount of data generated by excellent players to learn about the game in general.

Outside of Ultimate there is a huge amount of previous work in sports data. Tim follows the NBA heavily, and Ben is just interested in sports in general. We have obviously both been influenced by the reporting that goes on during games broadcasts and in the general media.

FiveThirtyEight in particular has done a lot of good statistics based reporting on sports, and produced a lot of good visuals. While they have some really out there stuff in terms of visualizations, their fall backs are bar graphs and scatter-plots. The amount of information they convey with these two basics leads us to believe we could don't have to jump straight into the deep end.

Grantland did some great reporting and made some cool visualizations that were less standard[11]. They would be a good place to look for inspiration after we've got the basics done.

We also did some research on what kinds of statistics were available and how they were presented during the 90's and early 2000's. This was a time that numbers were starting to gain traction, but still in their early stages. We referenced *Statistics in Sport* to

get a good idea of the landscape during this time. What we saw was that the statistics were still really basic, and the displays of them easily so. A lot of the discussion was of how to either generate or combine the basic statistics to make them more meaningful. We aim to mirror this. It would be better to create something that allows people to think critically about the information we have than to generate more information before we have done any thinking.

We also took some early inspiration from NBAVis, a site that does something similar to ours but for the NBA. It's less fleshed out with features, but still a good tool and it would be wrong not to mention them.

6 Research Question And Hypothesis

To state it explicitly: our research question is whether a tool could be created that would allow the ultimate community to interact with their data in a way they find meaningful. Our hypotheses is that people will feel that our tool allows them to generate and answer interesting questions, and get a better understanding of how well certain players perform.

7 Description

7.1 The Visual

For the visual it was important to us that we have as little clutter on the screen as possible. Where ever possible we did work to

either hide information when not needed, or do work behind the scenes. Because neither of us was that familiar with web development some of the data binding and actual dynamic stuff took a while, but we are happy with both the results and the skills we picked up.

The basics of the interface are the graph and the selection menu. The graph type can be changed via a drop down menu, and close to the axes are menus for selecting the stats. On the actual graph the labeling is minimal because we have enabled hovering over data points to get more information about them.

The selection menu is on the right. It contains the teams and players available to be selected. To add something to the graph find them here and click on them. Their name will be highlighted and the data automatically plotted. To remove them click on their name again.

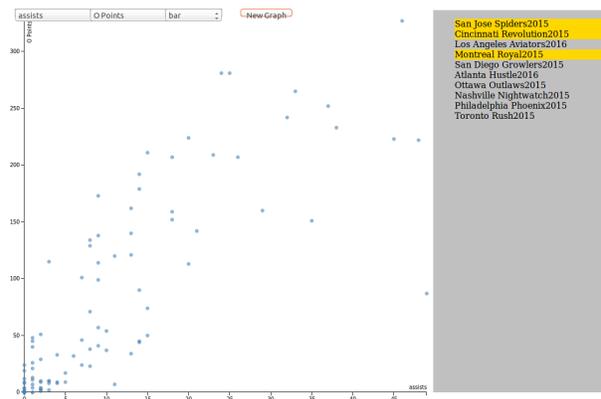


Figure 2: Here is our layout. Graph on the left, selection on the right, and menus up top

This interface may seem rather simple, but that was the whole point. We want anyone to be able to use and understand it. With as

basic a layout and as much work done ahead of time as possible we hope we have accomplished this.

Originally we wanted to color data points based on team colors, but we soon realized that team jerseys are too messy and their aren't yet defining colors attached to teams. Because of this we decided a two color scheme of blue and gold to avoid making people think they had found information where there is none.

We have had one problem with the bar graph when adding large amounts of data. As you add more data the bar's thickness becomes a small amount of pixels, and at certain amounts of data points it causes them to clump together in the middle. We haven't found a great solution to this, but making the graph bigger helps a lot.

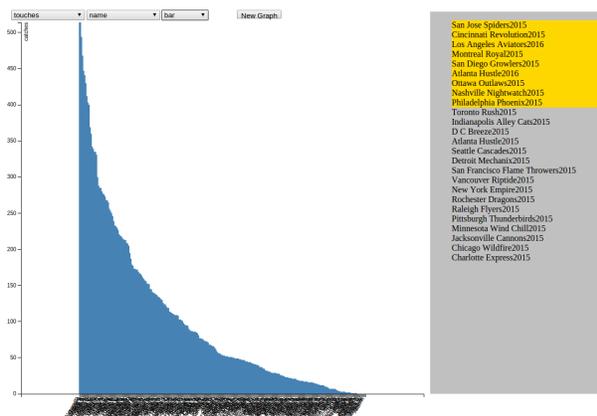


Figure 3: An example of the bars clumping together because they want to be next to each other but are only a couple pixels wide.

7.2 Creation

To create this visual we used D3.js. We owe a lot of credit to Mike Bostock and all of his extremely helpful tutorials and examples[6]. We are happy with the result we got from D3, but at many times it was a pain to use. We are developing on both Ubuntu and Mac OSX. Frequently something would work on only one of our machines, or only on Firefox but not Chrome. Not only that, but outside of the tutorials we found the documentation rather lacking. There wasn't an equivalent to man pages that we could find, which meant that every time we had a problem or wanted to use a function it was difficult to get actual specs.

As far as gathering the data, we explored a couple possibilities. The first was to rip the data from the HTML on the sites themselves. The AUDL's website, however, uses AngularJS to fill its pages with data after they load, which makes web scraping very tricky. We ended up downloading the play-by-play data for each team and tabulating this into the statistics we thought were important. This included pretty much every statistic we had ever seen tracked for Ultimate. We did this in python, and wrote the results to a results into a csv that we later read with javascript.

This tabulation took some work, as we had to write code for any stat that we wanted, and deal with a lot of edge cases in the data formatting. Now that we have an infrastructure set up though, it takes much less work to add stats as we need them so it is a good thing we went this way instead of going only with

the stats calculated by the AUDL.

While writing code to gather these statistics we created a visual debugging tool to show how successfully we were handling the files. A green line represents a success and a red line represents a failure.

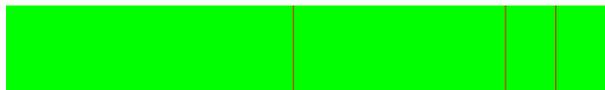


Figure 4: A visual of how well we processed a file. Green is good, red is a line we got nothing from. Most of the red lines are actually at cessations for validation of our tracking. These looked much more red earlier in development.

During the creation process we got some useful feedback from the class. They gave us some helpful tips on how to solve some of the problems we experienced with D3, as well as useful feedback on what features and layouts they considered the most important. Unfortunately, none of the classmates we talked to followed any sports so they didn't give us a lot of feedback on whether this was something our target audience might actually use.

8 Preliminary Insights

Though we have only had access to the tool for a small amount of time, and it is still undergoing development, we have already come to a couple interesting conclusions. We discuss these as an illustration of the tools power.

While looking at the number of touches each player got it quickly became apparent there was not the divide between handlers and cutters we might have expected. Handlers are sort of like the quarterbacks of ultimate - it is their job to handle the disc. Cutters are like the receivers - it is their job to generate yardage downfield. One might expect that by looking at the number of possessions that certain players on a team had you would be able to draw a line between the handlers and the cutters. And for some teams, this is true. But most of the time there is a smooth decline right down to the last man. And when one looks at all the teams at once there is a lack of any sudden drop-off anywhere.

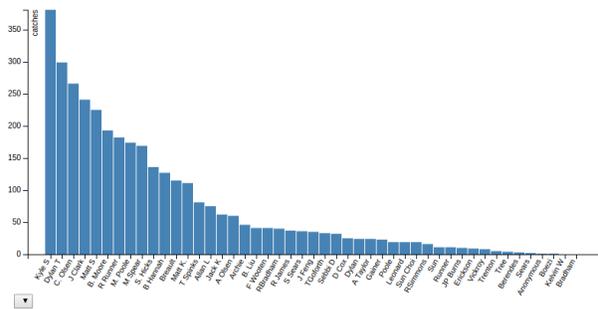


Figure 5: A plot of the number of touches for players on the Atlanta Hustle. Notice the smooth decline with no real cliffs.

Maybe the better statistic to look at would be touches/point. We don't have the ability to do this yet, but without this tool we never would have even asked the question.

Another thing we found that we wouldn't have otherwise was that the leading goal scorer was someone we had never heard of

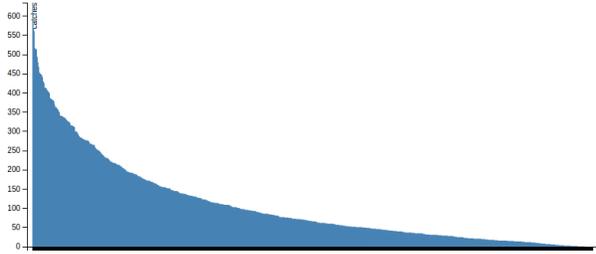


Figure 6: A plot of the number of touches for players across all of the AUDL. Again, notice the very smooth decline with no cliffs.

before (though the second most prolific was Beau Kittredge, widely regarded as one of the best in the game). Not only that, he accomplished the feat on a freakishly small number of touches. During the 2015 season Ethan Beardsley scored 70 goals on only 140 touches. This made him a pretty extreme outlier.

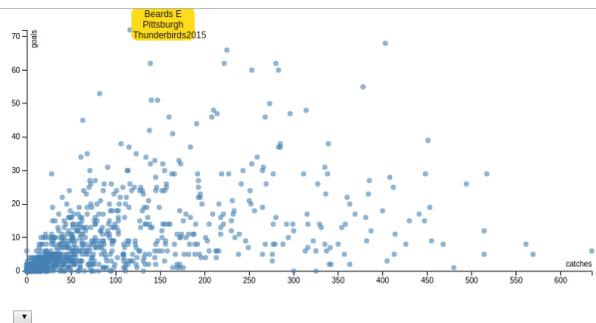


Figure 7: A scatter-plot of catches vs. goals. Far in the top left we see Ethan Beardsley, who scores pretty much every other time he touches the disc.

9 Bloopers

During our design process we definitely encountered some bugs, and some things we had not planned for. We've put some of them here.

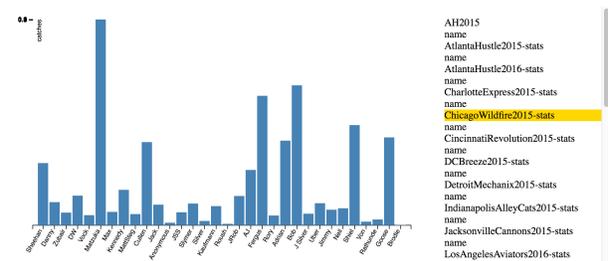


Figure 8: y-axis destruction. There are also no teams named "name". The bar chart is not sorted

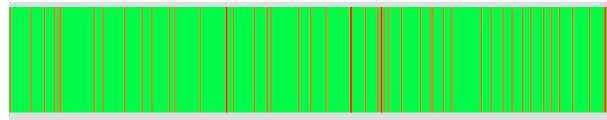


Figure 9: There were initially many lines of code that we failed to parse correctly.

10 Limitations

There are some definite limitations of our tool. The main one being that users are limited to only a couple basic visuals, and if they feel that a different graph would be better that is too bad. This may be less true in other tools, but the time taken to create a graph is also much longer. Another limitation is that currently we don't have it set

up such that users could easily feed in their own data. It could be done by someone technically savvy enough to run a python script, but that is a higher barrier to entry than we would like. This is something we could change by allowing people to upload stats, but there is a site that does that already (UltiAnalytics) that it doesn't make too much sense to compete with. Lastly, because of the fact that the original data is play by play users are limited to statistics that we have already calculated. We plan to expand this a little in future work, but are only able to go so far with it. With the statistics we do have there are some things we can't do that may be important. For example, we don't produce what the correlation between two stats is when looking at them on the plot. There are lot of things we could do here, but they may clutter the interface. We want to gather some user feedback before adding too many features. Lastly, there is a lot of data that simply isn't available yet. The easiest example is that it would be great if the play by play data included the location on the field, but it does not. There is no tracking of off-disc movement, and nothing like baseball has where they track ball velocity. If this were ever to be made available, our setup is not primed to visualize it well.

11 Results and Future Work

We haven't run a user study, but so far we are happy with the results. Both of us gen-

uinely enjoy using the app, and have used it to ask and answer questions we had not previously thought about. There are a couple features we plan to add in the immediate future, perhaps even by Tuesday. These include a search bar for players and a way to navigate to the stats page of an individual player by clicking at their data point on the graph. We would also like a way for players to craft their own statistics, e.g. touches/point. After we are done with this we plan to post our site on Reddit's Ultimate subreddit. Reddit is currently the main place people go on the internet to talk about Ultimate. From posting there we hope to get some exposure, and some feedback about what people thought and what features we should add. If the feedback was good, we would go about implementing them.

12 Division of Labor

In our project proposal we were docked some points for not being clear in the division of labor. However, about 90% of the time we spent working on this project we were sitting at the same table coding together and looking over each other's shoulders. As a result both of us were deeply involved in every part of the project, and we both did approximately the same amount of work.

References

- [1] Carl Bialik: Ultimate Frisbee Is In The Dark Ages Of Analytics - And It Wants To Escape *fivethirtyeight.com*.
- [2] Baseball Reference *<http://www.baseball-reference.com/>*.
- [3] ESPN *espn.go.com*.
- [4] Ulti Apps *ultiapps.com*.
- [5] USA Ultimate *<http://www.usultimate.org/index.html>*
- [6] Mike Bostock Personal Site *<https://bost.ocks.org/mike/>*
- [7] Fangraphs: What is WAR? *<http://www.fangraphs.com/library/misc/war/>*
- [8] Baseball Prospectus: VORP *<http://www.baseballprospectus.com/glossary/index.php?search=VORP>*
- [9] Jay Bennett *Statistics in Sport* Arnold Application of Statistics, 1998
- [10] NBAVIS *<http://www.cc.gatech.edu/gvu/ii/sportvis/nbaVis/vis/index.html>*
- [11] Grantland *<http://grantland.com/>*