Proceedings Paper:

Block, Florian Oliver orcid.org/0000-0003-0348-6731, Hodge, Victoria J. orcid.org/0000-0002-2469-0224, Hobson, Stephen John et al. (5 more authors) (2018) Narrative Bytes:Data-Driven Content Production in Esports. In: TVX 2018 - Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video. ACM , pp. 29-41.

https://doi.org/10.1145/3210825.3210833

Narrative Bytes: Data-Driven Content Production in Esports

Florian Block, Victoria Hodge, Stephen Hobson, Nick Sephton,
Sam Devlin, Marian F. Ursu, Anders Drachen, Peter I. Cowling
{firstname.lastname}@york.ac.uk,
Digital Creativity Labs, University of York, UK

ABSTRACT
Esports – video games played competitively that are broadcast to large audiences – are a rapidly growing new form of mainstream entertainment. Esports borrow from traditional TV, but are a qualitatively different genre, due to the high flexibility of content capture and availability of detailed gameplay data. Indeed, in esports, there is access to both real-time and historical data about any action taken in the virtual world. This aspect motivates the research presented here, the question asked being: can the information buried deep in such data, unavailable to the human eye, be unlocked and used to improve the live broadcast compilations of the events? In this paper, we present a large-scale case study of a production tool called *Echo*, which we developed in close collaboration with leading industry stakeholders. *Echo* uses live and historic match data to detect extraordinary player performances in the popular esport Dota 2, and dynamically translates interesting data points into audience-facing graphics. *Echo* was deployed at one of the largest yearly Dota 2 tournaments, which was watched by 25 million people. An analysis of 40 hours of video, over 46,000 live chat messages, and feedback of 98 audience members showed that *Echo* measurably affected the range and quality of storytelling, increased audience engagement, and invoked rich emotional response among viewers.

Author Keywords
Esports, Data-Driven Storytelling, Content Production.

CCS Concepts
Information systems → Information systems applications → Multimedia information systems → Multimedia content creation

INTRODUCTION
Esports is the term used for describing video games that are played competitively and watched by, normally large, audiences [34]. Over the past decade, esports have evolved from a niche segment of the games culture into a mainstream global phenomenon. In 2017, over 388 million people worldwide played or watched esports, and the number of esports fans is projected to grow a further 50% by 2020 [11].

ACM ISBN 978-1-4503-5115-7/18/06.
https://doi.org/10.1145/3210825.3210833

Esports take a variety of shapes, from digital versions of traditional sport, such as *FIFA* [4] to first-person shooters (e.g. *Counter Strike: Global Offensive* [16]) and magic-themed fantasy games, such as *Dota 2* [22] and *League of Legends* [15]. Many esports titles have associated amateur and professional online leagues and tournament, some of which feature prize pools exceeding those of many traditional sports [43, 52]. Coverage of esports borrows many elements from traditional sports broadcasting. Pre-game coverage features expert panels and follows the athletes as they enter the arena. During matches, virtual in-game footage is accompanied by audio commentary. After the match, coverage typically consists of interviews with the players and post-match analysis.

Esports also introduce a set of interesting changes to content production and delivery that sets them apart from traditional sports. In esports, live broadcast and on-demand content is almost exclusively delivered online, via video-platforms such as Twitch [18] and YouTube. Many esports titles also deliver live and on-demand matches as *raw data streams* that capture every aspect of the virtual world, such as the movement of the players. On the PC of a viewer, this data can be reassembled to create an interactive view of the match in which spectators can change camera angles, rewind, and interact with virtual objects. Match data also contains many statistics, timings and other additional layers of information that are not usually visible to the naked eye. Using specialised tools, such data can be extracted and translated into audience-facing content, such as graphs and statistics, enhancing the viewing experience.

In traditional sports, *tracking data* [44, 24, 46] has been used for a while to augment the viewers' experiences [75, 37], for instance, to better illustrate the performance of an athlete or make complex strategies visible to the viewer. In esports, the use of data and statistics have particular potential to benefit coverage and to help mainstream audiences enjoy and extract meaning from watching professional esports [34, 75, 37]. Compared to many traditional sports, gameplay in esports can be much more complex, making it hard for non-expert viewers to follow the action [34]. At the same time, as esports audiences grow rapidly, content creators increasingly need to make those complexities palatable and entertaining to casual audiences [13]. How can esports data be used by content producers to better engage their audience and make watching esports more meaningful to broader audiences?

In this paper, we present the first large-scale case study of *data-driven content production* in esports – i.e., the use of data analytics and data mining to inform the production of esports live coverage. We developed a production tool, called *Echo*, which uses large volumes of historic match data to detect highlights in live matches of the game Dota 2 [22], and provides mechanisms for automatically translating these highlights into audience-facing graphics. We deployed *Echo* at one of the largest international esports tournaments – ESL One Hamburg 2017, watched by 25 million people online – which allowed us to gather a wealth of experimental data consisting of: observational ethnographic data on how *Echo* impacted commentary and content production, materialised as 40 hours of video footage, 9 million chat messages and feedback from 98 directly surveyed audiences members, which we used to infer how *Echo* influenced the audiences' experience. We conclude that even simple graphical overlays of data-driven insights, such as the ones implemented by *Echo*, can have measurable effects on the commentary and quality of coverage. Many of *Echo's* graphics provoked elaborate discussions among commentators, and elicited strong emotional engagement among viewers. This is, indeed, the expected outcome, but, here, we provide strong experimental evidence coming from a significant case study – one of the largest esports tournaments in the world in whose production we deployed our *narrative byte tool*, *Echo*.

RELATED WORK

While the scientific relevance of esports has been highlighted over a decade ago [69], academic research in the area has been relatively sparse, with pockets of work ranging from economics [76, 63], to social practices [62, 26], regulation [36], and gender inequality [33, 41, 57]. Only a small number of studies directly address aspects of content production [25, 29, 65]. Studies in game-design have focused on optimising the visual presentation and user interface of esports games for observers [25, 29]. Work in machine learning has presented new mechanism for automatically extracting highlights from esports videos [65]. Outside of content production, most relevant to our work are studies of viewer needs and consumer practice in esports, technical work that leverages esports data for game analytics, as well as literature from traditional sports that inform our work. Each is presented in the following sections.

Consumer Needs and Practices in Esports

A series of prior work focuses on understanding viewer needs and consumer practices [34, 42, 72, 66], finding a variety of motivations for why people watch esports, including escapism, novelty and acquiring knowledge about the games being played. The latter motivation is particularly relevant for this paper, and translates into clear practical implications that make the case for data-driven content production esports. Hamari et al. [34] point out that content producers need to develop "better ways for the spectators to acquire knowledge from the eSport", explicitly pointing out the need to "more effective ways of displaying the game states", and highlighting the potential of overlay statistics. A comparison of consumption motives of esports and traditional sports found that one of the predominant motives for engaging with esports are competition and skill building [42], lending further importance to fostering skill acquisition as part of the viewing experience. In this paper, we expand on prior work by contributing the first survey of how data-driven content affects viewers' experience.

Esports Analytics

Esports data is often publicly available, and in terms of its detail and volume rivals even the most data-rich traditional sports, such as Formula One. The data-rich ecosystem surrounding esports has attracted attention by a series of work in the area of Game Analytics – the practice of extracting insights from data generated through digital games [32]. Esports are particularly conducive to the study of new techniques in Machine Learning and Artificial Intelligence, which require large volumes of training data to perform well. Examples of such application include identifying team encounters in esports data [59], predicting match outcome [38, 39, 35, 40, 47, 61, 64, 71], identifying factors that determine success in professional play [31, 55, 56, 74], as well as recommender systems that help players make better tactical decisions [23, 27]. Our work draws on similar analytical techniques. However, our focus is on studying the use of data and analytics as a *narrative tool*, which involves the translation of live data into formats that mainstream audiences can understand. Aside from academic work, we draw on an emerging landscape of esports data portals that seeks to make large volumes of esports data accessible. For instance, for the popular esport Dota 2, various data portals exist, such as Dotabuff [3], OpenDota [12] or DatDota [2], that essentially aggregate match data from tens of thousands of games and identify the most successful strategies over time. In line with the remit of our work, these platforms are targeted at mainstream audiences, and cater for the information appetite observed among the esports audience, as highlighted in the previous section. In this paper, we conceive a new category of applications that leverages data as a *real-time narrative tool for engaging live mainstream audiences*, with profound implication on the underlying design principles and implementation.

Data-Driven Content in Traditional Sports

Esports borrows many elements from traditional sports broadcast, making work relating to data-driven content relevant to the current investigation. From a commercial viewpoint, infographics and augmented views are part of most sports coverage today, such as Hawk-Eye Vision in tennis, visualisations of player's distance from the goal in football, or superimposing visual markers of current records in Olympic swimming competitions. Recent trends focus on the capture and visual representation of live sports tracking data, including 3D reconstructions, event and highlight detection, as well as ball and player tracking [28, 75]. Such techniques enable broadcasters to give their viewers new insights into the performance, strategy and style of athletes [75] as well as "explain things that weren't explicable previously" [37]. In this paper, we explore the foundations for data-driven content production in esports, with the aims of both opening up a new area of academic inquiry and contributing to the professionalisation of data-driven content production in the esports industry.

CASE STUDY

We conceived *data-driven content production* as the process of translating raw match data into audience-facing content. In the absence of existing commercial products, standards and prior academic work in this space, it was crucial that our research involved close collaboration with industry. Working with industry was key to acquiring an understanding of current practices, gathering requirements, and to co-design new technology [58]. Industry collaboration is also vital for creating opportunities for deploying and evaluating new tools in an ecologically valid setting, such as an esports tournament, involving expert crews that typically plan and execute such events.

Between June 2016 and March 2017 we ran a series of workshops and meetings, in which we recruited key stakeholders in the esports industry for participation in our co-design process. We recruited 17 senior experts across leading organisations within esports:

- *ESL* – Electronic Sports League [5] – is the globally leading esports company, producing events in two dozen countries and hosting some of the largest international esports tournaments, reaching tens of millions of viewers every year. Participating representatives from the company included several high level executives, senior product managers, and technical leads across TV, technology and pro gaming divisions.
- *Fnatic* [7] is one of the most successful esports organisations, having leading professional teams in a variety of esports, with players from all over the world. Participating representatives included senior executives as well as a former coach and data analyst.
- Prominent online celebrities that are regularly involved in some of the largest international esports events, including a leading commentator, two celebrity analysts, and two casters (individuals who broadcast their own games online), with a combined total of over 2 million online followers.

The research team working with the domain stakeholders consisted of experts in interactive storytelling, user-centred design, interactive data visualisation, game analytics, and machine learning.

Requirements Gathering

To inform the design of a concrete prototype, we wanted to obtain insight into any existing practices of using data during the production of live esports events. Our initial set of workshops provided the following insights:

Data = USP (Unique Selling Point). A consensus across representatives from big esports companies and freelance talents was that esports data has significant potential to augment and advance current broadcasting practice. Large content producers see data-driven content as a potential means of generating USP that will more deeply engage the tech-savvy fan base, and potentially make esports more attractive to watch for new viewers. Individual content producers echoed this sentiment, highlighting that innovations in augmenting in-game footage could help them individualise their content.

Data is already used, but clunky to work with. Various members of our expert team, in particularly our stage analysts and commentators, routinely use historic data for storytelling, mostly taken from existing data portals, self-made tools and pre-assembled databases. However, it was lamented that drawing on data involves significant manual labour. Many existing tools only provide very basic ways of probing and visualising data, while more sophisticated tools are hard to operate. Moreover, most tools do not provide flexible options for presenting the data in audience-friendly way, making them unsuitable for real-time storytelling.

Crews are small, data expertise is rare. Esports productions have very small profit margins. Compared to regular sports broadcasts, crews are extremely streamlined and data expertise is rare. It is desirable that new data-driven production tools can be operated by existing production staff.

Simplicity is key. Esports broadcast are already overloaded with information (see for instance Figure 1). Any additional messaging has to be simple enough to be consumed and understood by the audience in seconds.

Feedback by our domain experts provided a clear set of design requirements. Tools for data-driven content production in esports need to: 1) provide a high level of automation in data acquisition and processing, 2) integrate mechanisms that help operators identify interesting patterns within complex data, and 3) support a 'one-click' mechanism for translating data into extremely simple graphics that audiences can instantly understand.

Context – Defence of the Ancients 2

Having elicited basic design requirements we needed to choose a specific esports game to contextualise our case study. In consultation with our experts, we chose the game *Dota 2* [22]. Not only is Dota 2 one of the most popular esports [49], it also provided the most comprehensive public access to both live and historic match data. Dota 2 also features rich and complex gameplay, with high potential for data-driven content production to enhance the viewer's experience. Lastly and somewhat more pragmatically, our pool of experts included a high-profile caster, analysts and commentator specialising on Dota 2, giving us access to game-specific domain-knowledge during the design process.

Dota 2 is a fantasy-themed game in which two teams of five players compete (think 5-a-side football with various weapons and endowed with magic). Each player picks from a pool of 115 distinct *hero* avatars (or characters), each

Figure 1. In-game view of Dota 2 as seen by the players and the audience (© Valve Corporation).

having unique skills and play styles. At the start of the game, heroes find themselves in their own base (one base per team) located at opposite corners of a square-shaped virtual arena, consisting of forests, three roads and a river. The overall aim of the game is for teams to defend their own base while destroying the enemy base. To fulfil this objective, players need to first strengthen their heroes, by collecting two basic types of in-game currencies: gold and experience. Gold can be used to purchase virtual items that make heroes more powerful. Experience (XP) allows heroes to level up, increasing their overall strength. Gold and XP are obtained by killing enemy heroes, or computer-controlled monsters called *creeps* that are scattered across the arena. In their effort to compete for resources and conquer each other's base, teams often clash. When taking too much damage, heroes can "die", rewarding the enemy team both gold and XP, as well as sending the dead hero back to its base. Killing enemy heroes allows teams to temporarily outnumber the enemy team, creating further tactical advantage and opportunities to break down the enemy's defence. Matches have no predetermined length, but typically last anywhere from 20 minutes to over 1 hour.

It's All About the Player

Our experts stressed that many key stories of Dota 2 broadcasts gravitate around "the player" (analogous to traditional sports [75]). However, tracking each player's performance, even for expert viewers and commentators, is often challenging. Unlike the strongly simplified account given above about Dota 2's game mechanic, gameplay in Dota 2 is extremely complex, with players performing various parallel objectives while mostly being dispersed across the map. During a live match, commentators often resort to a set of basic statistics built into the game that captures each player's performance according to a set of *Key Performance Indicators* (KPIs). One of the most important KPIs, for instance, is a player's *Net Worth* - the total gold value the hero has acquired. Seeing an ordered list of Net Worth gives some basic understanding on which player is doing well at the current time. However, an issue is that Net Worth is only indicative of good performance for some heroes. Other heroes have different objectives, such as helping their team-mates or scouting portions of the terrain that is controlled by the enemy. However, these objectives are not captured by out-of-the box KPIs.

Another issue with the game's built-in statistics is that they only reflect data from the current game. Relative comparison between different heroes, however, is often not meaningful due to vast differences in playstyle. Some heroes for instance, are strong early into the game, while others start off weak and get stronger later in the game. In order to fully judge how well each player is performing with their chosen hero, one has to rely on previous experience of observing the same hero being played. Not every viewer, however, has this knowledge for all 115 heroes. Some broadcasts have used historic comparison to better illustrate live performance to audiences. For instance, Ben Steenhuisen, Dota 2 analyst and developer of DatDota [2], cross-references historic databases with real-time data during a game, and creates small text snippets illustrating players' performances that are shown to the audience in form of a small popups. This process involves a set of exclusive expert tools, significant technical skill, expertise in data science, and manual labour and to be operated effectively. Another limitation of the process is that the way information is presented to the audience cannot be customised. Reflecting our design requirements, we wanted to expand existing practices by creating a production tool that is easy to operate, assists with identifying extraordinary data points based on historic data, provides an entirely automated data pipeline, and allows free customisation of audience-facing visuals.

Echo – Capturing the Extraordinary

To improve the audience's ability to track hero performance, we developed a new tool called *Echo* (see Figure 2). *Echo* is a production tool that can monitor data from a live match and, for each minute into the game, compare performances of each player to thousands of historic performances. *Echo* displays one column per player, five per team. Within each column, *Echo* shows a list of 12 KPIs that capture various aspects of the player's performance, significantly expanding Dota 2's out-of-the-box statistics. The KPIs were chosen as a result of technical availability (what we could access in the live data stream) and our expert's advice for covering a range of different playstyles and roles. *Echo* calculates the percentiles of the live performances – or how many percent of historic performances are exceeded in the live game. For each KPI and player, *Echo* provides three values: 1) The first percentile expresses how the player's performance at the current minute relates to thousands of previous performances of the same hero. For instance, the magnified portion of Figure 2 shows that 42 minutes into the live match, the player named "inYourdreaM" has reached a Net Worth that is higher than 98% of values previously recorded of all players that chose the same hero. 2) The second percentage shows how the player's live performance compares to the player's personal match history. This column is suitable to detect if a player has a particularly good or bad game compared to their historic performance. 3) The third column shows how the current performance is situated within *all* matches within the reference data, including performances by any player and any hero. This column is particularly suited to track "all

Figure 2. Basic view of *Echo*, showing each Player's live performances as percentiles of historic data.

Figure 3. Detailed view of a particular KPI. If "Show to audience" button is pressed, *Echo* triggers the graphics system to show a graphical overlay to the audience (© ESL).

time" records or extraordinary performances across heroes. Values that exceed any value in the scope are marked with "REC" = record. Negative records (worst values ever observed in each respective scope) are marked as "LOW". Values that exceed certain thresholds are colour coded (yellow = record, purple = top 1%, blue = top 5%, red = negative record). The colour coding assists the operator with identifying interesting constellations. For instance, in the row Networth highlighted by the magnified portion of Figure 2 one can deduce that for the hero, the player lies within the top 2% of historic performances, but does relatively poorly (bottom 16%) compared to their personal match history. This indicates that this player usually picks heroes that are capable of obtaining gold faster. The third column tells us that in the global reference frame, taking into account all heroes and players, the player performs only above average.

Clicking any percentile opens a more detailed information dialog, giving the operator additional information for the specific performance, including a histogram, current minimum and maximum records and a preview of the type of core messaging that would be shown to the audience (see Figure 3, left). Pressing the button in the top right corner of the preview will automatically generate an audience-facing graphic, which can then be superimposed over the in-game footage (see Figure 3, right).

The main goal of the design of the graphics was to reduce complexity and to provide a single, clear message to the audience. As stories revolve around players, a natural part of the visual design would be a portrait, and the player's *gamer tag* (the in-game nickname). The more difficult task was to wrap the core message into a clear presentation. As it is well established, special care needs to be taken when conveying statistics to the public, as statistics literacy may widely differ and even basic statistical concepts, such as percentiles, may not be well understood [51]. Consequently, we presented the percentile with a qualitative cue "Top X%" or "Bottom X%", including an adjusted percentage that creates a correct representation of the percentile. For instance, In Figure 3, *Echo* would have identified the player's performance as being in the 97th percentile, translating into the phrase "Top 3%". While we did not expect most viewers to read information beyond the core messaging, we did want to provide a second level of information that commentators and quick readers could access, essentially describing the scope of the performance. The percentiles depend on the current minute in the game as well if one compares current performance to all games of a certain hero, the player's match history, or all games in the database. To visually break up this information, we added two sets of information to the graphics. Above the core message, we displayed the current value of the KPI (e.g. "X gold") and the time the graphics referred to ("@ minute X"). Below the graphics, we added small print that described the scope (hero, player, or all games) as well as the range of games that are considered in our database. As part of our evaluation, we will assess how those different components were perceived.

Implementation

Echo was implemented in C# and WPF, utilising a local SQLite database [17] for storing the historic match data. The historic match data was compiled to include all professional games for the current season, including a total of 13,550 performance metrics (across 1,355 matches, each containing 10 player performances). Raw match data was obtained using the official Dota 2 match history API [20] in conjunction with APIs from Datdota's match API [2] to obtain a url for the raw replay file (storing the full recording of the match). We then downloaded all 1,355 raw replay files and processed them using a customised version of Clarity [16], an open-source match parser for Dota 2. The parser allows to traverse match data in a chronological way, and extract time series of all KPIs per player.

EVALUATION

Echo was deployed at ESL One Hamburg 2017, one of the largest Dota 2 tournaments of the year, which featured a $1m prize pool and attracted over 25 million online viewers as well as 20,000 fans onsite [6]. With a schedule stretching across four days, the event provided an ideal environment to collect rich observational data about how *Echo* impacted the viewers' experience, and to gather explicit audience feedback on *Echo* as well as the general importance of data-driven content in Dota 2.

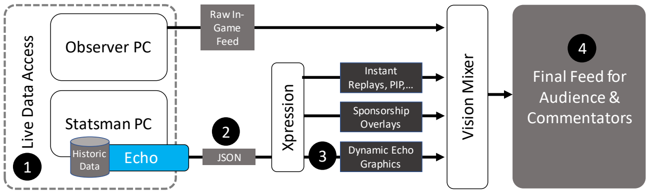

Figure 4. Schematic of Echo's integration: (1) Echo accesses real-time data through the Statsman's PC, (2) sends selected data points to Xpression via JSON http request; (3) Xpression generates a graphic and overlays the final feed (4).

Integration with Production Workflow

Esports coverage is structured very much like traditional sports coverage. A pre-game panel introduces the match, giving background on teams and players, followed by the players entering the arena and taking their gaming station on stage. This sequence is produced in line with traditional broadcast. The actual in-game coverage, is produced by a separate team, most notably the audience-facing talents – usually two commentators. Behind the scenes, the primary roles for the production of in-game coverage during an esports tournament are the *statsman*, the *spotter* and the *graphics operator*. The statsman is responsible for technically setting up matches and monitoring statistics during the match. The spotter is the virtual camera operator. The statsman and spotter work on independent PCs, which both run Dota 2's observer mode, allowing them to watch the players in real-time, control camera angles and navigate across the virtual environment, and access to the built-in statistics tools. The spotter's in-game view is the main footage going into the vision mixer. Analogous to traditional sports broadcast, the graphics operator is responsible for enriching the raw in-game footage with visual overlays, such as sponsorship information, instant replays, or picture-in-picture segments. The mentioned observer mode in Dota 2 also provides a real-time data API, which is the entry point for *Echo*, which was installed on the statsman's PC. *Echo*'s dashboard was displayed on a secondary monitor. *Echo* was integrated with the graphics software used by the production [12] (see Figure 5), enabling the statsman to bypass the graphics operator and directly push graphics to the audience.

The preparation of the graphics-integration, including visual and animation design was carried out by ESL, to reflect their visual design language used throughout the tournament. The statsman had full editorial control over timing and selection of what performances from *Echo*'s data dashboard he wanted to show to the audience. As a general practice, graphics were shown during quiet moments of the match, to not distract viewers from important events. As *Echo*'s graphics always related to an individual player, the statsman frequently coordinated with the spotter, so that the in-game view focused on the player who the graphics related to.

Data Collection

The four-day tournament provided various opportunities for collecting observational data and gathering audience feedback. The unedited live footage, as seen by the audience, was made freely available on Youtube after each day, capturing every occurrence of *Echo's* graphics and how it affected commentators' narrative – our first data source. The majority of viewers watched the event coverage live, via Twitch [18], which provided them with a real-time chat in which they could communicate with each other during the broadcast. We recorded all these time-stamped messages for analysis. Between matches, we recruited onsite viewers for in-person interviews. On the fourth event day, we also conducted an online survey, and recruited participants via social media. In total, we collected 40 hours of broadcast footage, 46,000+ chat posts from twitch, 69 online audience survey responses and 29 in depth interviews with visitors of the event. We also collected user activity logs of *Echo*, recording which graphics the operator showed to viewers, when they were shown and for how long. This allowed us to accurately time-align the display of *Echo*-generated graphics with both the audience chat log and the actual tournament footage.

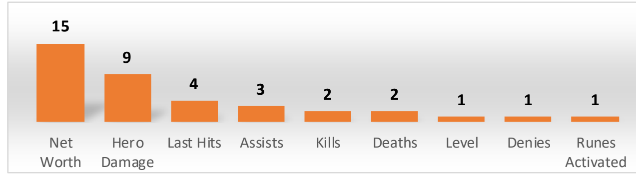

Figure 5. Frequency of graphics by KPI

Summary of Use

Across 27 matches played at ESL One Hamburg, *Echo* generated 38 audience-facing graphics (average 1.4 / match). In the majority of matches (21), *Echo* produced 1 or two graphics. In two matches, *Echo* did not produce any graphics, due to a lack of extraordinary stats. In three matches, three graphics were shown, while in one game, four graphics aired. The average air time per graphic was 10 seconds, totalling in 7.5 minutes of airtime received during over the course of the whole tournament. Of all graphics shown, 27 graphics (71%) illustrated *positive* performances, such a player performing in the top 5% or breaking a top record. 11 graphics (29%) highlighted *negative* performances, such as poor records.

Figure 5 shows the frequency of graphics by specific Key Performance Indicators (KPIs). This distribution yields various insights. The most frequently used KPI – Net Worth – is widely acknowledged to be the most important statistic used by experts to gauge hero performance, so it is unsurprising that it became a popular choice for the data operator. The second most frequently used KPI – damage done to enemy heroes – is not usually exposed through the game client. This provides good indication that *Echo* was effective in expanding the built-in statistics, and that this information was deemed relevant to the audience by an experienced statsman. There was also a good breadth of KPIs used at the tail end of the distribution – while the two most frequently used KPIs make up for exactly 50% of the alerts, the other 50% are distributed across 7 different KPIs. This shows that *Echo*'s vocabulary of KPIs generate narratives that brought a more varied aspects of performance to the fore.

Impact of *Echo* on Commentary

Based on review of video footage showing the occurrence of all 38 graphics generated by *Echo*, we evaluated the qualitative effect the graphics had on the commentators' narrative. Overall, out of the 38 total graphics, 14 of the graphics elicited commentators to make direct verbal reference to the contents of the graphic. In six cases the graphics were aired in direct response to the commentators' narrative. In four instances, both effects interleaved, creating a dialog between commentators and statsman. Following an ethnographic approach of data analysis [70, 67], the following paragraphs aim to highlight the *quality* of different effects observed. We do not claim representativeness or general validity of our findings. Rather, we wanted to identify and characterise an initial set of categories of impact that can inform future theory and experimental study in the area of data-driven content production.

Shaping Narrative, Eliciting Surprise. In various instances, the graphics highlighted an aspect of a player the commentators had not been aware of, or apparently had not considered. When a graphic about a *Net Worth Record* was displayed (placing the player at the top 1%), the commentators noted:

Commentator 1: "Look at that….look at that"
Commentator 2: "Yeah, he's really fat [high networth]"

The first statement captures a degree of surprise, indicating that the commentators had not been aware of this extraordinary performance. The graphics triggered an elaborate subsequent discussion about the player's strategy and performance. In another instance, *Echo* generated a graphic highlighting that a player was performing in the *Top 2% of Assists*. Assists are the number of enemies killed in presence of this player. This graphic indicated that the player's involvement in team-fights was very high. This triggered the following consideration by the commentator:

Commentator 1: "Here we can see his early levels may have been a little slow, in terms of experience, but just his involvement, his movements… always where the action is"

Again, the statement provides evidence of a degree of surprise, moving focus from a KPI in which the player was underperforming, to a KPI that brought out a different aspect of the player's contribution that had thus far gone unnoticed. Overall, we observed 8 instances in which the reaction of commentators indicated a certain element of surprise. In some instances, *Echo*'s graphics shaped the commentators narrative minutes after they appeared.

Doing Poorly: highlight bad performances. Eleven graphics pointed out poor performances, sometimes eliciting strong reactions by the commentators. One graphic pointed out the very poor level of a player named "Solo" (levels indicate progress of a player) eliciting an emotional response by the commentator:

"I mean, Solo [=player's name], bless him…" [continued]… he's had a lot of deaths…", and it's the levels, as you pointed out earlier, really holding him back. Level 13, 43 minutes in."

The initial comment was followed by an elaborate discussion between commentators about the reasons for this poor performance, bringing out various insights to the audience. Stats also brought out comic moments. 16 minutes into a match, a certain player had only killed a single creep (this is roughly equivalent to a tennis player not having scored a single point in a whole set). The graphics, marked a "poor record", invoked amusement among the commentators:

Commentator 1: "Look at this stat man." [Both chuckling]
Commentator 2: "Poor record, one last hit – bless him."

Data-Dialogs. In four cases, we observed clear dialogs between commentators and statsman. In one instance, the commentators are talking about how well the hero named "Rubick" is doing:

Commentator 1: "Look at the money on Rubick!"
Commentator 2: "This is gotta be one of the richest Rubicks I've seen in a long time."

While commentator 2 finishes the sentence, an *Echo* graphic pops up showing that Rubick is breaking the Net Worth record (= how "rich" a hero is), prompting an immediate reaction by the other commentator:

Commentator 1: "Top record for net worth – 21 thousand on him at the 61 minute mark […]"

In another match, the graphics slots in perfectly with the commentators narrative about a hero named "Chen" performing very well in terms of his Net Worth:

Commentator 1: "This Chen is getting very rich…" [Graphic popping up showing that the hero "Chen" is within the top 5% of Net Worth] "…statistically rich."

Overall, the graphics generated by *Echo* stimulated vivid discussions between the commentators, producing informative and entertaining background facts to the audience. Data-dialogues were observed, in which statsman and commentators engaged interactively to shape a coherent narrative. There was clear evidence that *Echo* brought aspects to the fore that surprised commentators and diversified the usual narrative toolbox of statistics.

Impact of *Echo* on Twitch Chat

To analyse the impact of *Echo*'s graphics on the audiences' engagement on Twitch chat, we applied a combination of quantitative and qualitative measures on the twitch chat data. *Frequency of Posting* has been proposed as a quantitative proxy measure for engagement [53]. For each graphics shown on the stream, we calculated frequency of posting 30 seconds before and after the graphics appeared. A repeated measure factorial design (ANOVA) with two factors *Timing* (before / after) and *Graphics Type* (positive performance / negative) showed a significant effect for *Timing* ($F_{1,33}$=11.616, p=.002). On average, frequency of posting rose from a baseline of 3.7 contributions just before the appearance of the graphic, to 4.9 after the graphic aired. There was also a significant interaction between *Graphics Type* and *Timing* ($F_{1,33}$=17.448, p=.027), showing that the increase in chat activity was significantly more pronounced for negative performances (87.1% increase in frequency of posting) compared to positive ones (12.5% increase).

To get a qualitative understanding of how *Echo*'s graphics affected chat engagement, we conducted an analysis of all 30 second chat excerpts, comprised of 5091 chat posts. The following paragraphs describe our findings:

Repetition, Repetition. A substantial portions of chat entries replicate either the value of the KPI (e.g. "95 net worth"), or the percentage (e.g. top 5%). Among the 5091 analysed chat posts, a total of 1,242 (~20%) contained one or more repetition of the graphics content (487 repetitions for content of "top performances", 755 repetitions for content of "bottom performances"). This means that a substantial percentage of overall chat contribution focused on re-sharing the messaging of the graphics on chat.

Rich emotions. To characterise the makeup of the other 80% of chat messages, as well as to characterize the conversational context in which the graphics contents were shared, we conducted a word frequency analysis. Of the top 20 most frequently used words (accumulatively making up almost half of all word occurrences), 10 words were abbreviations, such as "LOL" (=laughing out loud), or "Clap". The other ten words were iconic in nature. Chat users commonly express their emotions explicitly in form of memes or emoticons, which we interpreted using a variety of online sources [1,9,19]. For instance, when typed in twitch chat, the word *PogChamp* is represented with the icon  (head of a person with a wide open mouth), which is commonly used to express shock and disbelief [1]. The observed memes reflected a range of emotions, including surprise, tension, frustration, upset and amusement.

Effectiveness of Visual Design

The presented findings so far suggest that the design of *Echo*'s graphics was effective in bringing out the central message to the audience (e.g. "Top 5%"). Across commentary and Twitch chat, there was also evidence that other details had been picked up. For instance, after a player broke a top record, the following dialog ensued:

Commentator 1: "Hey look, here we go. Top Networth"
Commentator 2: "Top record."
Commentator 1: "At 8 minutes."
Commentator 2: "8 minutes."
Commentator 1: "I mean [mumbles] it's crazy. We are pre 9 minutes and we are seeing heroes reaching […]"

In this and three other instances, commentators picked up on the timing aspect contained in smaller letters above the main message, and utilized this information from the graphics to emphasise the core message. However, commentators never picked up on the "small print" below the main message, and there was also no indication that commentators picked up the difference between scopes. In one instance, this led to the commentators portraying a performance as global record, even though the graphics depicted only a personal record (referencing only a single player's match history). A handful of Twitch users did, however, pick up on the small text, expressing concern with the phrase *"since July 2017"*, with some users expressing confusion or disagreement: "SINCE JULY 7 MAN NOT ALL TIME WTF[=outrage]", "top 5% damage since july?". While there were only three total instances among 5091 chat posts, this evidence suggests that a small number of users do scrutinise even "the small print", and raise questions of the validity of the stats. We only include matches since July, since then a major patch (update that changes the game) had significantly impacted various KPIs, making it questionable to compare data collected before and after the patch. An alternative wording "since patch 7.07" may have addressed those concerns.

As for visual language, evidence showed that we were successful in contextualising players performance in respect to historic data, and current time of the match. However, we were not successful in bringing out differences between the different scopes, or reference frames, that *Echo* provides. In future iterations, this information should be more clearly visible, possibly even integrated with the core message (e.g. "PLAYER TOP 5%" vs. "HERO TOP 5").

Online Survey

To elicit explicit feedback on how the graphics were received we used an online questionnaire. The two aims of the online questionnaire were a) to identify generally what role data-driven content plays in viewer's experience of Dota 2 broadcasts, and b) to capture specifically how they rated the graphics generated by Echo. Participants were recruited through advertisement via Twitter, which was re-tweeted by the official ESL account and the tournaments hashtag #ESLOne. The following sections discuss the main findings. 106 people responded to the survey, with 69 people fulfilling the criteria of being over 18. 63 participants fully completed the survey. Figure 6 shows all questions and the results.

The majority of respondents (70%) felt that statistics were 'very important' or 'extremely important' to their experience of Dota 2 (Q1). When asked about how important statistics are when *watching* a Dota 2 broadcast, 81% of subjects responded with 'very important' or 'extremely important' (Q2). In addition, another 16% of subjects chose 'moderately important'. Notably, none of the subjects responded with 'not at all important'. After the two general questions, subjects were shown a sample graphic generated by Echo, followed by four questions specifically referring to the sample graphic. The majority of respondents (76%) leaned towards finding *Echo*'s graphics easy to understand, where as 15% leaned towards the opposite (Q3). 10% did not feel strongly either way. 56% did not find *Echo*'s graphics distracting, opposed by 27% of participants who did (16% neutral, Q4). The majority of subjects (84%) found that the graphics contained useful information. When asked about whether subjects wanted to see more of this information in the future, 84% subjects responded with 'strongly' agree (55%) and 'somewhat agree' (29%).

The results capture an interesting design trade-off between distracting audiences and delivering useful value-add information. Data from the online survey provided evidence that overall, *Echo* was successful in its core aim of bringing useful and meaningful information to the majority of respondents, even to some of those who perceived Echo as distracting. The responses also captured high demand and appetite among viewers for data-driven content in Dota 2.

Semi-Structured Interviews

To complement the quantitative findings from the online survey, we conducted 29 semi-structured interviews with fans in the arena to shed light on *the possible qualitative aspects* of data and statistics for the purpose of storytelling. In line with our analysis of commentary, our focus was not on representativeness or general validity. Instead, we wanted to extend prior statistical findings in consumer demand by providing a more facetted picture of potential motives for the like or dislike of data-driven content. Individuals or groups were randomly approached to participate in the study. Audio of all interviews were recorded. The questions were:

- Q1. What role do statistics play in your experience of Dota 2?
- Q2. What role do statistics play as part of broadcasts such as ESL ONE?

- showing subjects a sample graphic generated by Echo -

- Q3. Did you notice those graphics during the event?
- Q4. What did you think of them?

Broadly, answers to Q1 and Q2 echoed results from the online survey. 25 interviewees of the 29 made references to statistics having an important role in their experience of Dota 2 as players and viewers. As theorised by prior studies of consumer demand in esports [34], our data confirms that statistics in Dota 2 help both novice and experienced viewers *better understand the current state of the game*, and keep track of the many intricacies of professional play: *"[statistics are] important because you can't keep track of everything"*. Along the same lines, a novice viewer remarked: *"For me it's totally interesting to get a better overview of the game"*.

One group of participants expressed that statistics let them better predict likely outcomes: *"[live statistics] help us to guess on what will happen in the future"*. This points to the potential of statistics to foster the audience's *anticipation,* which, in turn, sets the stage for viewers to be surprised, e.g. when an anticipated outcome does not occur.

Using *Echo* as an exemplar, Q3 and Q4 brought out additional nuances of how data-driven content can enrich a live event. Concretising advantages of "keeping track", interviewees expressed that *Echo* helped them get *"a better look at the players and how they are doing"*. Another

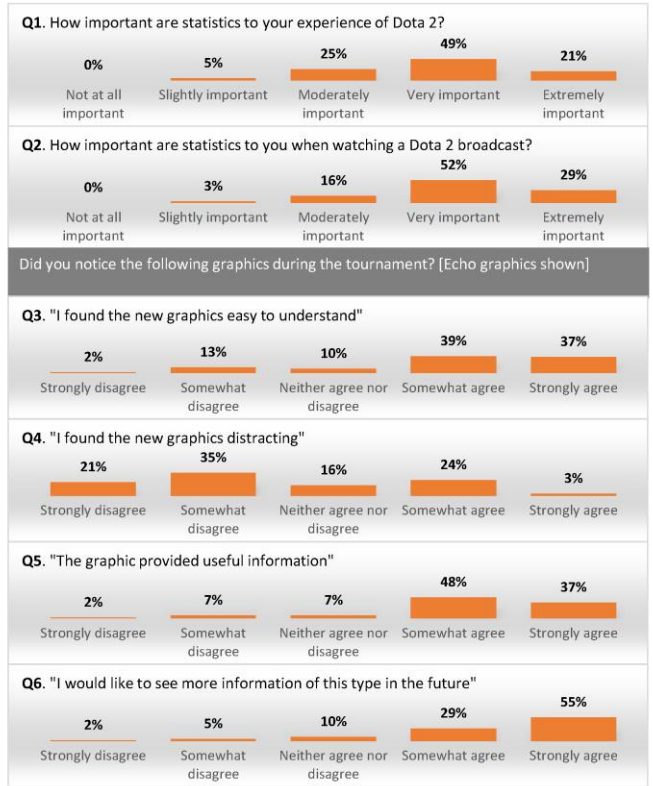

Figure 6. Results from the online Survey, see text for details.

participant felt *Echo*'s graphics helped enhance his appreciation for professional play: "*"I think it's quite impressive to see what level they are reaching in the playing skills"*. Entertainment was also brought out as a quality of *Echo*'s graphics: *"It's fun! Yeah, I like it"*. Particularly the negative performances provoked some explicit mention: *"Particularly when watching the ESL One, you can really see when people really fail at the game […] it's funny"*.

DISCUSSION & CONCLUSION

Echo provides substantial evidence for the importance of data-driven content production in esports. Our audience survey shows that data and statistics are central to the viewing experience, and audiences have a clear appetite for more data-driven content, such as is presented by Echo. Through co-designing *Echo* with leading industry partners, and conducting a large-scale observation of how it was used in one of the world's largest esports tournaments – we generated a range of detailed insights that extend existing research on consumer motives in esports and traditional sports [34, 37, 42, 72, 66, 75]. We provide observational evidence confirming that data-driven content can indeed be an effective tool to make the gameplay more transparent to viewers [34], to cater for the audience's desire for skill building [42], as well as to enable audiences to more effectively follow the performance of the athletes [75].

Our data paints a more facetted picture of consumer demand for data-driven content in esports. Data-driven content not only allows content producers to *"explain things that weren't* *explicable previously"* [37], but also brings to light aspects of gameplay that often go unnoticed, such as lack of engagement of specific players, unusual strategies and tactics, and behaviour which is measurably different than what has gone before (particularly record-breaking). This demonstrates the potential of using data to extend the breadth of storytelling. Echo helped audiences contextualise performances within historic data, and factors of the game (e.g. how much time has passed). Beyond illustrating the current state of the game, audience feedback also suggests that data-driven content may be helpful to foster anticipation and hence increase attendance at future events. Many esports neither have a fixed match length, nor a clear mechanism that shows the current "score" – making it hard for audiences to know what to expect. The ability to anticipate, however, seems crucial to setting the stage for an element of surprise, which in turn, characterises many memorable sports event – anticipation and surprise being crucial ingredients of good storytelling. This suggests that prediction algorithms, originally designed for betting and game analytics [38, 39, 40, 47, 61, 64, 71], could have useful application in storytelling, e.g. showing audiences a simple display of win percentage that helps them judge likely outcomes. *Echo* has demonstrated that even simple graphical overlays of data-driven insights can have measurable effects on the quality of commentary and coverage. Many of *Echo*'s graphics provoked elaborate discussions among commentators, and invoked strong emotional engagement among viewers.

It is important to note that the presented findings are generated in the concrete context of Dota 2 and do not necessarily generalise to other esports games. However, Dota 2 has almost identical game mechanics to League of Legends [15], another representatives of the MOBA genre (Massive Online Battle Arena). Dota 2 and League of Legends make up two of the three most popular esports [49], having an estimated combined following of 113 million people [73]. Arguably, this makes our findings relevant for a substantial portion of the ecosystem in esports content production. Further studies will have to be conducted to study the relevance of data-driven storytelling in other popular genres, such as first person shooters. Another area of future study is how we can translate insights from esports, which is data-rich today, to the broadcast of traditional sports and other live events that may in the foreseeable future have access to data-streams of similar quantity and quality.

While *Echo* currently generates static content, in our future work we will explore how data can engage audiences in a more interactive fashion. For instance, companion apps could offer a personalised experience of data-driven content that adapts to the needs and interests of individual viewers. In contrast to traditional sports, gameplay data in esports not only exists for professional matches, but for *all* matches. This would enable forms of engagement that were not possible in traditional broadcast, for instance, by comparing professional performance with each individual viewer's personal performance to give individualised insights and coaching tips.

The high volumes of esports data also provides a unique environment to study the role of Machine Learning and Artificial Intelligence as creative tools for storytelling. In our future work, we will explore the use of more sophisticated AI to help content creators identify unique moments in gameplay, and to provide an AI assistant making it possible to generate personalised, interactive data-stories. We also aim to create data-driven tools that help small production crews and individual streamers create rich coverage.

Our case study has also demonstrated that esports provides a unique ecosystem for capturing audience feedback and behavior. Each esports event inherently generates a rich set of digital observational data, much of which can be automatically collected and analysed. While still more men watch and play esports than women, the esports audience is racially diverse [13], and much more inclusive in regards to disabilities than traditional sports [68]. This makes esports a fertile environment for experimentation, for creative production, social enquiry, and cultural analysis. Our findings highlight the appetite of esports fans to engage with data as a creative tool – involving not only large production companies, but millions of individual content producers that could share data-driven stories peer-to-peer. In many ways, one could conceive data-driven storytelling as a tool for millions of esports fans to connect with each other creatively. From an educational perspective, this opens up a set of attractive possibilities for study. Esports constitutes a unique environment to study how mainstream audiences wish to engage with and extract meaning from data. An unexplored educational benefit of content-driven storytelling in esports may thus be its ability to foster data literacy and "mental fitness" by motivating vast young audiences to learn through telling stories with esports data about which they care deeply.

ACKNOWLEDGEMENTS
We want to thank ESL for their support and providing access to ESL One Hamburg. Special thanks go to Marcel Menge, Tobias Grieser, Christian Eikermann and James Lampkin for their time and help during the design, deployment and evaluation of Echo. We also want to thank Ben "Noxville" Steenhuisen and Jonathan "JJ Pimpmuckl" Liebig for their careful and creative operation of Echo during ESL One Hamburg, as well as to Ted "Pyrion Flax" Forsythe, Kevin "Purge" Godec and Tobi "TobiWan" Dawson for sharing their advice and expertise during the design of Echo. We also thank Fnatic's Patrik "cArn" Sättermon, as well as Luis "Deilor" Sevilla and Fabien Ungerer for their thoughtful input. Last but not least, our gratitude goes to the members of the ESL One Hamburg audience who have taken part in the study. This work was conducted at Digital Creativity Labs (www.digitalcreativity.ac.uk), supported by EPSRC / AHRC / InnovateUK under grant no EP/M023265/1.

REFERENCES

1. A beginner's guide to the most-used Twitch emotes: https://www.dailydot.com/unclick/twitch-emotes/
2. Datdota. https://www.datdota.com/
3. Dotabuff. https://www.dotabuff.com/
4. EA. FIFA Video Games. https://www.ea.com/games/fifa
5. ESL, https://www.eslgaming.com/
6. ESL, Press Release ESL One Hamburg, 2017. https://www.eslgaming.com/press/esl-one-hamburg-powered-intel-debuts-20000-fans-onsite-and-25-million-online-viewers
7. Fnatic. https://www.fnatic.com/
8. Java Script Object Notation, ECMA-404 The JSON Data Interchange Standard. https://www.json.org/
9. Know your meme http://knowyourmeme.com/memes/monkas
10. Newzoo. Esports revenues will reach $696 million this year and grow to $1.5 billion by 2020 as brand investment doubles (2017). https://newzoo.com/insights/articles/esports-revenues-will-reach-696-million-in-2017/
11. Newzoo. Global Esports Market Report (2017). https://newzoo.com/insights/trend-reports/global-esports-market-report-2017-light/
12. OpenDota. https://www.opendota.com/
13. PWC. The burgeoning evolution of eSports (2016). https://www.pwc.se/sv/pdf-reports/the-burgeoning-evolution-of-esports.pdf
14. Ross, Real-Time Motion Graphics, Xpression. https://www.rossvideo.com/graphics-system/xpression/
15. Riot Games, League of Legends. http://leagueoflegends.com/
16. Skadistats. Clarity Parser for Dota 2. https://github.com/skadistats/clarity
17. SQLite Consortium, SQLite. https://www.sqlite.org/
18. Twitch. https://www.twitch.tv/
19. Twitch-A-Speak http://mashable.com/2014/08/08/twitch-emoticons/#pQdQuLh3HGqc
20. Valve, Dota 2 Match History WebAPI. https://dev.dota2.com/showthread.php?t=47115 (accessed 30/01/2018)
21. Valve, Counter-Strike Global Offensive. http://blog.counter-strike.net/
22. Valve. Dota 2.http://www.dota2.com/play/
23. Agarwala, A. and Pearce, M., 2014. Learning DotA 2 team compositions. Technical report, Stanford University.
24. Alamar, B. and Mehrotra, V. (2012, February). Beyond Moneyball: The Future of sports analytics. Analytics Magazine. Retrieved from: http://www.analytics-magazine.org/special-articles/525-beyondmoneyball-the-future-of-sports-analytics
25. Carlsson, C. and Pelling, A., Designing Spectator Interfaces for Competitive Video Games. Master's thesis, Report No. 2015:129, Chalmer University of Technology, Gothenburg, Sweden 2015.
26. Carter, M. and Gibbs, M.R., 2013, May. eSports in EVE Online: Skullduggery, fair play and acceptability in an unbounded competition. In FDG (pp. 47-54).
27. Conley, K., Perry, D., 2013. How does he saw me? A recommendation engine for picking heroes in DotA 2. Technical report, Stanford University.
28. Dietrich, C., Koop, D., Vo, H.T. and Silva, C.T., 2014, October. Baseball4d: A tool for baseball game reconstruction & visualization. In Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on (pp. 23-32). IEEE.
29. Ditmarsch, van, J.L., 2013. Video games as a spectator sport (Master's thesis). Report no 3268969, Utrecht University, 2013.
30. Dove, G. & Jones, S. (2012). Narrative Visualization: Sharing Insights into Complex Data. Paper presented at the Interfaces and Human Computer Interaction (IHCI 2012), 21 - 23 Jul 2012, Lisbon, Portugal.
31. Eggert, C., Herrlich, M., Smeddinck, J. and Malaka, R., 2015. Classification of player roles in the team-based multi-player game DotA 2. In: International Conference on Entertainment Computing. Springer, pp. 112-125.
32. El-Nasr, M.S., Drachen, A. and Canossa, A., 2016. Game analytics. Springer London Limited.
33. Fox, J., & Tang, W. Y. (2016). Women's experiences with general and sexual harassment in online video games: Rumination, organizational responsiveness, withdrawal, and coping strategies. new media & society, 1461444816635778.
34. Hamari, J. and Sjöblom, M., 2017. What is eSports and why do people watch it?. Internet research, 27(2), pp.211-232.
35. Hodge, V., Devlin, S., Sephton, N., Block, F., Drachen, A., & Cowling, P. (2017). Win Prediction in Esports: Mixed-Rank Match Prediction in Multi-player Online Battle Arena Games. *arXiv preprint arXiv:1711.06498*.
36. Hollist, K.E., 2015. Time to be grown-ups about video gaming: the rising eSports industry and the need for regulation. Ariz. L. Rev., 57, p.823.
37. Horky, T., & Pelka, P. (2017). Data Visualisation in Sports Journalism: Opportunities and challenges of

data-driven journalism in German football. Digital Journalism, 5(5), 587-606.

38. Johansson, F., Wikström, J., 2015. Result prediction by mining replays in DotA 2. Master's thesis, Blekinge Institute of Technology, Karlskrona, Sweden.
39. Kalyanaraman, K., 2015. To win or not to win? A prediction model to determine the outcome of a DotA 2 match. Technical report, University of California San Diego.
40. Kinkade, N., Jolla, L. and Lim, K., 2015. Dota 2 win prediction. Technical report, University of California San Diego.
41. Kim, S. (2017) Gender inequality in eSports participation : examining League of Legends . Unpublished Masters Thesis, Univeristy of Texas at Austin. http://hdl.handle.net/2152/62914
42. Lee, D. and Schoenstedt, L.J., 2011. Comparison of eSports and traditional sports consumption motives. ICHPER-SD Journal of Research, 6(2), pp.39-44.
43. Lewis, L. and Bradshaw, T. 2017. Esports: Is the gaming business ready to come of age? Financial Times, Noveber 5th, 2017. URL: https://www.ft.com/content/ef8539b6-be2a-11e7-9836-b25f8adaa111
44. Lewis, M. (2004). Moneyball: The Art of Winning an Unfair Game. W. W. Norton and Company.
45. Lim, Y.K., Stolterman, E. and Tenenberg, J., 2008. The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. ACM Transactions on Computer-Human Interaction (TOCHI), 15(2), p.7.
46. Lindsey, G. R. (1959). Statistical Data Useful for the Operation of a Baseball Team. Operations Research, 7(2), 197-207.
47. Makarov, I., Savostyanov, D., Litvyakov, B. and Ignatov, D. I., 2018. Predicting Winning Team and Probabilistic Ratings in DotA 2 and Counter-Strike: Global Offensive Video Games. Springer International Publishing, Cham, pp. 183-196.
48. Medlock, M. et al. The Rapid Iterative Test and Evaluation Method: Better Products in Less Time. Cost Justifying Usability, An Update for the Internet Age. Boston, Morgan Kaufmann, 2005.
49. Meola, A. Here are the most popular eSport games of 2017 and the companies behind them. Business Insider, Dec 2017. http://uk.businessinsider.com/top-esports-games-developers-2017-11
50. Millington, B. and Millington, R. (2015).'The Datafication of Everything': Toward a Sociology of Sport and Big Data, Sociology of Sport Journal 2015 32:2, 140-160
51. Murray, S. and Gal, I., 2002, July. Preparing for diversity in statistics literacy: Institutional and educational implications. In Proceedings of the Sixth International Conference on Teaching of Statistics. Ciudad del Cabo: IASE.
52. Nordmark, S. The top 10 highest prize pools in esports, The OP, 24th December 2017. URL: https://dotesports.com/the-op/news/biggest-prize-pools-esports-14605#list-1
53. Pan, R., Bartram, L. and Neustaedter, C. TwitchViz: A Visualization Tool for Twitch Chatrooms. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16). ACM, New York, NY, USA, 1959-1965.
54. Pingali, G., Opalach, A., Jean, Y. and Carlbom, I., 2001, October. Visualization of sports using motion trajectories: providing insights into performance, style, and strategy. In Proceedings of the conference on Visualization'01 (pp. 75-82). IEEE Computer Society.
55. Pobiedina, N., Neidhardt, J., Calatrava Moreno, M. d. C. and Werthner, H., 2013. Ranking factors of team success. In: Proceedings of the 22nd International Conference on World Wide Web. ACM, pp. 1185-1194.
56. Rioult, F., Métetivier, J.-P., Helleu, B., Scelles, N. and Durand, C., 2014. Mining tracks of competitive video games. AASRI Procedia 8, 82-87.
57. Salter, A., & Blodgett, B. (2012). Hypermasculinity & dickwolves: The contentious role of women in the new gaming public. Journal of broadcasting & electronic media, 56(3), 401-416.
58. Sanders, E.B.N. and Stappers, P.J., 2008. Co-creation and the new landscapes of design. Co-design, 4(1), pp.5-18.
59. Schubert, M., Drachen, A. and Mahlmann, T., 2016. Esports Analytics Through Encounter Detection Other Sports.
60. Segel, E. and Heer, J., 2010. Narrative visualization: Telling stories with data. IEEE transactions on visualization and computer graphics, 16(6), pp.1139-1148.
61. Semenov, A., Romov, P., Korolev, S., Yashkov, D., and Neklyudov, K. (2016). Performance of Machine Learning Algorithms in Predicting Game Outcome from Drafts in DotA 2. In International Conference on Analysis of Images, Social Networks and Texts (pp. 26-37). Springer, Cham.
62. Seo, Y. and Jung, S.U., 2016. Beyond solitary play in computer games: The social practices of eSports. Journal of Consumer Culture, 16(3), pp.635-655.
63. Seo, Y., 2016. Professionalized consumption and identity transformations in the field of eSports. Journal of Business Research, 69(1), pp.264-272.

64. Song, K., Zhang, T. and Ma, C., 2015. Predicting the winning side of DotA 2. Technical report, Stanford University.

65. Song, Y., 2016. Real-time video highlights for yahoo esports. arXiv preprint arXiv:1611.08780.

66. Sjöblom, M. and Hamari, J., 2017. Why do people watch others play video games? An empirical study on the motivations of Twitch users. Computers in Human Behavior, 75, pp.985-996.

67. Spradley, J.P., 2016. Participant observation. Waveland Press.

68. Usmani, B. This is Your Brain on Esports, Motherboard, Vice, 2014. https://motherboard.vice.com/en_us/article/mgbe9x/this-is-your-brain-on-esports

69. Wagner, M.G., 2006, June. On the Scientific Relevance of eSports. In International Conference on Internet Computing (pp. 437-442).

70. Walsh, D., 1998. Doing ethnography. Researching society and culture, pp.217-232.

71. Wang, W., 2016. Predicting multiplayer online battle arena (MOBA) game outcome based on hero draft data. Master's thesis, Dublin, National College of Ireland.

72. Weiss, T. and Schiele, S., 2013. Virtual worlds in competitive contexts: Analyzing eSports consumer needs. Electronic Markets, 23(4), pp.307-316.

73. Wolmarans, Kyle. Dota 2 vs. League of Legends: Updating the numbers. Criticalhit, 2016. https://www.criticalhit.net/gaming/dota-2-vs-league-legends-updating-numbers/

74. Yang, P., Harrison, B. E. and Roberts, D. L., 2014. Identifying patterns in combat that are predictive of success in MOBA games. In: Proceedings Foundations of Digital Games.

75. Yu, X. and Farin, D., 2005, July. Current and emerging topics in sports video processing. In Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on (pp. 526-529). IEEE.

76. Zarrabi, S. A., & Jerkrot, H. N (2016). Value creation and appropriation in the esports industry. Department of Technology Management and Economics Division of Innovation Engineering and Management, CHALMERS UNIVERSITY OF TECHNOLOGY, Gothenburg, Sweden. Report No. E 2016:090