A Brief Overview of Data Mining and Analytics in Games

Günter Wallner, Eindhoven University of Technology & University of Applied Arts Vienna
guenter.wallner@jku.at

Author Draft

Citation: Wallner, G. (2019). A brief overview of data mining and analytics in games. In G. Wallner (Ed.) *Data analytics applications in gaming and entertainment*, (pp. 1-14). Auerbach Publications. URL: https://www.taylorfrancis.com/chapters/edit/10.1201/9780429286490-1/brief-overview-data-mining-analytics-games-g%C3%BCnter-wallner

1. Introduction

The 21st century has been repeatedly proclaimed to be the century of data. The increasing processing capabilities of computers and the proliferation of Internet-enabled devices have made it easier than ever to gather more and more data about every aspect of our daily lives. The massive amount of data produced every day, however, also needs to be transformed into actual information and knowledge to be of real value and to be of actual use for decision making. This requires adequate techniques and tools that help uncover hidden and valuable information or patterns within the collected data. Otherwise, the large volumes of data are just that – data bare any deeper meaning. This is the goal of data mining (Kantardzic, 2011). While data mining has been and is receiving considerable attention to be able to cope with the ever-increasing data volumes, the term has started to appear in the late 1980's, early 1990's (Coenen, 2011; Dong & Pei, 2007). Data mining is not a single technique but rather a conglomerate of methods, techniques, and algorithms, usually applied in an iterative or explorative process.

The 21st century is, however, not only considered to be the age of data mining but has also been coined the ludic century by game designer Eric Zimmerman (2015) – an age that is characterized by play. It may thus only be fitting that during the last decade or so, data mining has found its way into game production and has become a crucial part of game development and maintenance. This has led to the emergence of the new field of 'game analytics' – broadly speaking, the application of analytics to game development and research (Drachen, El-Nasr, & Canossa, 2013). It is *"the practice of analyzing recorded game information to facilitate future design decisions"* (Medler, 2009, p. 188). Game analytics uses data mining techniques to discover patterns and to extract information from game-related data, especially player behavioral data. As it is often the case with new fields, the establishment of game analytics can hardly be tied to a specific point in time or be ascribed to a single factor. Instead, the emergence of game data mining and analytics may be rather attributed to a coincidence of several developments.

The first steps into the direction of game data mining have presumably been made at the turn of the century when online games such as *EverQuest* (Sony Online Entertainment, 1999) slowly started to track data about gameplay (cf. Weber, 2018). A couple of years later, in 2003, one of the first articles on how to improve game design through data mining was published (Kennerly, 2003). Although specifically focused on massive multiplayer online games such as the aforementioned *EverQuest*, many of the described techniques also applied to other types of games. Again a couple of years later, first articles in popular media appeared with a Wired article (Thompson, 2007) on how Microsoft relied on scientific methods to inform games user research being one of the earlier examples. However, it was not until about 2010 when data mining and analytics really started to gain momentum (Weber, 2018). In 2008, Microsoft Game Studios published one of the first research articles (Kim et al., 2008) on how tracking user behavior in games can contribute greatly to the design of video games. As of 2009, Medler (2009) attested that it is hard to find a digital game that does not allow to record gameplay in some way or the other but also noted that analyzing the recorded information to inform design is still in its infancy. A year later, Zoeller (2010) presented the telemetry suite BioWare is using for analyzing tracked behavioral data of players and Schoenblum (2010) presented the data collection backend developed at Epic Games. From then on, things happened very quickly and by now game analytics has become prevalent across industry and a major aspect of games research. It is an area which has seen substantial growth in the last ten years and which is still evolving rapidly.

This growing interest in data mining and analytics has been spurred by several developments and technical advances. First, the wide adoption of Internet-enabled gaming devices allows developers nowadays to remotely and unobtrusively track the behavior of a large numbers of players. Before that, playtesting usually happened by bringing customers in-house and observing them in a laboratory-style setting while they are playing the game. Consequently, this happened at a much smaller scale and the invited players may not have been representative of the whole player population of the game. However, as games have become a mainstream phenomenon and are being played by an increasingly diverse audience, it has become a matter of particular interest to create games that appeal to a wide range of players. In this sense, data mining can be a valuable tool for acquiring representative data. The possibilities offered by the Internet and modern mobile devices as well as advances in web technology have also paved the way for new types of games such as massively multiplayer online games or social network games played on social media platforms such as Facebook which, in turn, attract new audiences. These games are played by hundreds to thousands of players simultaneously and who may even interact with each other. This complexity makes such games challenging to develop, requiring extensive testing with a large player base to properly balance the game, to ensure a satisfying player experience, and to resolve and avoid technical issues. Remote data collection offers a natural and convenient way to gather such large-scale and long-term datasets. Moreover, production budgets of video games have risen considerably in the past years with budgets of ten to hundreds of million dollars not being uncommon anymore. For example, *Grand Theft Auto V* (Rockstar North, 2013) had an estimated development budget of $137.5 million (Sinclair, 2013) and development costs for *Gran Turismo 5* (Polyphony Digital, 2010) were reported to be $60 million (Remo, 2009). Even if these are extreme examples and not all budgets are this high, the required investments pose a great financial risk for developers in case a game fails. Through gathering actual in-game data, developers have a means to meet audience expectations and, in turn, achieve financial success. The ever-increasing production budgets also caused developers to find ways to extend the lifespan of games and to search for new business models to alleviate the associated risks. Among these are subscription-based services, downloadable content, micro-transactions (purchasing of virtual goods for a very small amount of money), or free-to-play games (games that are basically free to play with monetization happening through micro-transactions). Some developers have started to view 'games-as-a-service' rather than as a one-time purchase. A recent report from DFC Intelligence (Cole, 2018) suggests that the growth of EA and Activision, two of the biggest publishers in the industry, can to a large extent by attributed to this service model. In such a model revenue is also generated after the initial release using subscriptions or, for instance, by providing new content – spread over a longer period of time – in order to uphold audience interest. Retention of players is essential for such business models and data mining and analytics offers a valuable approach to monitor and study player engagement. All these developments have been fueled by advances in data storage and processing capabilities (Coenen, 2011) which allow to efficiently analyze the large volumes of data as they appear in game development today.

2. Applications

Consequently, data mining and analytics has been applied to a variety of purposes within game production and research. In the following four broad and common application areas are briefly discussed.

Data mining to improve design and player experience: Game development is a highly creative process that optimally needs to undergo continuous and critical evaluation to ensure that the final game is engaging and offers a satisfying player experience. This is the primary goal of games user research (GUR), which aims to *"help game designers reach their design goals by applying scientific and UX [User Experience] design principles, and by understanding players"* (IGDA GRUX, 2018). As data mining, GUR is not a single technique but rather a collection of qualitative and quantitative methods, such as playtesting (Fullerton, 2004; Mirza-Babaei, Moosajee, & Drenikow, 2016), biometrics (Nacke, 2015), interviews (Bromley, 2018), and surveys (Brühlmann & Mekler, 2018). Over the years, analytics has become a valuable addition and by now constitutes an essential component of GUR (cf. El-Nasr, Desurvire, Aghabeigi, & Drachen, 2013). Analytics offers many benefits for complementing existing methodologies as telemetry data promises a large-scale and objective view on player behavior (i.e., the data is not biased by players' subjective opinions) which would be difficult, or even impossible, to obtain through other methods. Unsurprisingly, data analytics has thus found broad application in GUR so far, reaching from developing behavioral profiles of player activity (Drachen, Thurau, Sifa, & Bauckhage, 2013) over the study of virtual economies (Castronova et al., 2009; Morrison & Fontenla, 2013) to all aspects of balancing, such as extracting reoccurring behavior patterns to detect dominant strategies (Bosc, Kaytoue, Raïssi, & Boulicaut, 2013; Wallner, 2015). Apart from that, there is also a large body of work focusing on spatial and spatio-temporal aspects of gameplay (Wallner & Kriglstein, 2012; Kang, Kim, Park, & Kim, 2013; Drachen et al., 2014), which is of particular importance as movement forms one of the most important mechanics in nearly all games. Analytics may also be used in combination with qualitative and observational GUR methods (Desurvire & El-Nasr, 2013) in order to provide context to each other, although triangulating the different data sources is not straightforward (Mirza-Babaei, Wallner, McAllister, & Nacke, 2014).

Data mining to inform business decisions: As discussed previously, the game industry is actively seeking alternative business models to the traditional pay-once format to reach new costumers. These models, such as free-to-play or subscription-based services, rely on keeping customers engaged over extended time periods and on providing attractive spending opportunities to generate revenue. At the same time, the number of games being released each year, and thus the number of games competing for customers, steadily increased. For instance, as of December 17, 2018 the statistics site SteamSpy counts 4,696 games being released on the digital distribution platform Steam[1] in 2016, while already 7,047 games were released in 2017 and 8,882 in 2018 (Galyonkin, 2017). Inevitably, it has become more challenging to acquire new customers and to stand out from the plethora of games already on the market. With the business becoming fiercer, analytics-based solutions provide ample opportunities to support business decisions. Analytics can provide insights into player retention and churn, such as how long players keep playing or where they are quitting (Bauckhage et al., 2012; Hadiji et al., 2014; Xie, Devlin, Kudenko, & Cowling, 2015). It can help answer questions concerning conversion rates, that is, what makes a player of free-to-play games convert into a paying customer (Hanner & Zarnekow, 2015; Fields & Cotton, 2011) and about the purchasing behavior of players, for example, for which in-game content players are willing to pay and why (Lehdonvirta, 2009; Hamari et al., 2017). In addition, analytics is vital for the prediction of customer lifetime value (the total amount of revenue earned from a player) (Chen, Guitart, del Río, & Periáñez, 2018; Sifa et al., 2015) and for customer acquisition, for instance, to help plan marketing campaigns (Williams, 2015a). Moreover, players also exert influence on each other which should not be underestimated as a game's community can contribute greatly to the success or failure of a game. In this sense, analytics can help to build stronger communities (see also below) and to improve community management (Williams, 2015b). These and other related possibilities are testament of the value and potential of data analytics for business intelligence in game development.

Data mining to innovate and optimize game technology: Data mining has become a cornerstone in the development and advancement of game technology. Indirectly, this application area may also yield better player experience as discussed previously but the main focus here should be on how data can help build technology. Among others, behavioral data forms an integral aspect for many artificial intelligence (AI) algorithms which take advantage of the massive amounts of domain knowledge contained within game-log data. Examples in this space include works on opponent modeling for AI bots such as the ones by Weber and Mateas (2009) who applied machine learning techniques to large collections of replay data in order to learn strategies and by Synnaeve and Bessiere (2011) who presented a Bayesian model focused on predicting the opening strategy of opponents – with the model parameters being learned as above from a set of labeled replays. Others, in turn, used tracked player data to drive content creation and to adapt games to the player. Player models, that is, models that encompass the player's traits and

[1] `http://store.steampowered.com/` (Accessed: December 2018)

behavior are a core component of these works and can greatly benefit from being data-driven (Hooshyar, Yousefi, & Lim, 2018). For instance, Pedersen, Togelius, and Yannakakis (2010) relied on tracked user data for level generation in a platform game whereas Missura and Gärtner (2009) as well as Zook and Riedl (2012) proposed dynamic difficulty adjustment algorithms based on collected player data. Data mining can also be used to build technology for supporting playtesting. Recent and seminal work by Stahlke and Mirza-Babaei (2018) explored how AI can assist in the automatic playtesting of games to reduce time and costs associated with usertesting. While this work primarily relied on geometry-based navigation[2] to mimic human behavior, it can be envisioned that by incorporating historic player data such approaches will become even more useful in the future as they will be able to more accurately imitate user behavior.

Data mining to empower players and to foster community building: While this area is probably not the first to come to mind, and may also be less present when discussing data analytics in games, it nevertheless constitutes an important and growing field of application. This is on the one hand reflected in the number of games that offer visualizations that allow players to inspect their recorded gameplay data (Bowman, Elmqvist, & Jankun-Kelly, 2012; Medler, 2011). Indeed, players have been identified as an important target audience for gameplay visualizations (Wallner & Kriglstein, 2013; Bowman et al., 2012). Hazzard (2014) identified two main purposes of such visualizations: representations to convey the in-game status and visualizations for training. While the former have traditionally focused on the active player, these visualizations have – in light of the increasing popularity of esports and of game streaming platforms such as Twitch – also become more and more common for observers. Training visualizations are less common within games itself but rather exist external to games. Their development has been spurred by game developers providing access to the collected data through public APIs as it is the case, for example, for *League of Legends* (Riot Games, 2018) and *Destiny* (Bungie, 2018). Other games, such as *Starcraft* (Blizzard Entertainment, 1998) or *World of Tanks* (Wargaming, 2010), record the gameplay in so-called replay files which can then be shared and analyzed. This has led to proactive efforts on behalf of the player community which is eagerly utilizing these data sources to build data-driven websites and tools. In that sense, the data serves as a driver for community building and prolonged involvement with a game.

The above discussed application areas are not to be viewed as distinct and independent from each other as, for instance, improving game design can impact player retention while new technologies can contribute to better game experiences. Nor should they be regarded as a complete account of applications of game data mining. Rather, they should serve as examples for the immense possibilities game analytics offers.

[2] See Algfoor, Sunar, and Kolivand (2015) for an overview of navigation modeling in games.

3. Limitations

While data mining and analytics have taken on an important role in game development, creating new and exciting opportunities and offering new benefits, it should not be seen as a panacea for all problems. For example, one of the primary strengths of telemetry data – being objective and unbiased, that is, not being influenced by subjective perception and reporting (Kennerly, 2003; Drachen & Canossa, 2009; Wallner, 2013) – is also a kind of limitation as it does not provide reliable data about 'why' players behaved the way they did (Kim et al., 2008; Lynn, 2012). Moreover, analytics is not always a straightforward a trivial process, thus making critical reflection and careful interpretation of the data a necessity to avoid pitfalls and to derive meaningful results. These issues may range from issues with the data itself such as low data quality and unrepresentative data to problems associated with data interpretation such as 'cherry picking' results that best support a certain hypothesis or being susceptible to confirmation bias, i.e., seeking of information that supports a pre-existing hypothesis (see, e.g., Thomson, Lebiere, & Bennati, 2014). Data analytics requires expertise and experience to accurately model player behavior (e.g., Powell, 2016) and to draw the right conclusions.

With the shift towards quantitative data collection, and particularly with the advent of free-to-play games (cf. Takahashi, 2017), the notion of data-driven game design – where decisions are largely based upon the collected data (King, Churchill, & Tan, 2017) – has emerged. In an effort to alleviate risks potentially caused by making choices based on intuition, developers may increasingly resort to more objective metrics. This has raised critical voices that an over-reliance on quantitative analytics bears the risk of replacing individual creativity by a – what Whitson (2012) coined – 'design by numbers'. Or to put it even more drastically, there is a fear that the art of game design is progressively marginalized and instead replaced by scientific techniques. Others, in turn, have raised concerns that games may become more homogenized in order to appeal to a large audience (see, e.g., Whitson, 2012; psychotrip, 2016). These are reasonable doubts if data becomes the sole basis for decision-making. Game developers, however, repeatedly highlight the importance of analytics being only a resource to empower designers but not the replace them by analytics (Carr, 2015; Mansell, 2015; Koskenvoima & Mäntymäki, 2015). Instead, data mining and analytics should be regarded as a powerful toolset to inform decisions. In a data-informed design, data is one but not the only input on which decisions are made (King, Churchill, & Tan, 2017).

4. Visual Analytics

As already hinted at above, two frequently named challenges (see, e.g., Powell, 2016; Koskenvoima & Mäntymäki, 2015) associated with game data mining are that it is 1) complex and requires sufficient expertise and skill, and 2) that communicating the results so that they are easily understood and acted upon is not always straightforward but at the same time essential. These two are, however, not specific to game analytics but rather apply to data analytics in general. Information visualization has been recognized (e.g., Keim, 2002) as a powerful tool to assist with the analysis process and can help with analysis, both confirmative and explorative, and presentation (Keim, Mansmann, & Schneidewind, 2006). The aim of confirmatory analysis is to – as the name already implies – the confirm or (or falsify) an a-priori formed hypothesis. Exploratory data analysis (EDA), in contrast, does not seek to answer specific pre-existing hypotheses but rather to discover patterns, trends, or anomalies. EDA can be very useful to develop an initial understanding of the data, to form or refine hypotheses, and to identify new directions for the analysis. As games are complex systems which can give rise to emergent behavior hard to anticipate beforehand, EDA takes on a critical role in game analytics (Wallner & Kriglstein, 2015).

Information visualization is beneficial for both approaches as it takes advantage of the cognitive and perceptual abilities of humans (Fekete, van Wijk, Stasko, & North, 2008). Indeed, many argue (e.g., Shneiderman, 2002; Keim, 2002) that the analysis process is most effective if automatic analysis techniques are combined with interactive visualizations allowing for more efficient reasoning and decision-making. This integration of the processing and analytical capabilities of computers and humans' perceptual abilities is an integral part of 'Visual Analytics' (cf. Keim et al., 2008). The benefits of both, information visualization and visual analytics, have been recognized early on among game analysts (Kim et al., 2008; Zoeller, 2010) and form nowadays an indispensable part of game data mining and analytics (Wallner & Kriglstein, 2013). This includes the use and adaption of existing visualization techniques such as heatmaps and node-link diagrams (e.g., Thompson, 2007; Canossa, 2009; Andersen, Liu, Apter, Boucher-Genesse, & Popović, 2010; Wallner, 2013) for gameplay analysis. One of the earliest examples of gameplay visualization is the work of Hoobler, Humphreys, and Agrawala (2004) on visualizing player behavior patterns in competitive team games. Since then many visualizations tools and algorithms dedicated to gameplay analysis have been developed. Examples include, DataCracker (Medler, John, & Lane, 2011) a visual game analytics tool build at Electronic Arts, Ubisoft's DNA suite (Dankoff, 2014) which offers various visualization and data exploration capabilities, G-Player (Canossa, Nguyen, & El-Nasr, 2016) a visualization system focused on spatio-temporal data, and PLATO (Wallner & Kriglstein, 2014) which integrates several visualization techniques and analytics methods such as clustering and subgraph matching.

While the target audience of the aforementioned examples are first and foremost developers, visualizations must not be restricted to developers but can also be specifically targeted towards players (Bowman et al., 2012; Wallner & Kriglstein, 2013). For instance, Wallner (2018) proposed an algorithm which automatically creates 'battle maps' from recorded combat data to give players a means to retrospectively reflect on their performance. In a similar vein, Kuan, Wang, and Chuang (2017) described a visualization system for data from real-time strategy games to help players learn new strategies. Indeed, with more and more games also providing public access to the collected data, the player community has started to use this data to create visualizations on their own (e.g., Belicza, 2014; Temmerman, 2017).

5. Conclusions

To conclude, data mining and analytics have paved the way for new innovations in game development and for the evaluation of player behavior. As shortly outlined in this chapter, game analytics has already found many applications including design, technology, business, and community relations but it can be anticipated that the role of analytics will continue to increase in the future. With the 21st century being considered both the 'ludic century' and the era of big data, it is only fitting that we are currently witnessing the fusion of big data and games.

References

Algfoor, Z. A., Sunar, M. S., & Kolivand, H. (2015). A comprehensive study on pathfinding techniques for robotics and video games. *International Journal of Computer Games Technology*, *7*.

Andersen, E., Liu, Y. E., Apter, E., Boucher-Genesse, F., & Popović, Z. (2010). Gameplay analysis through state projection. In *Proceedings of the 5th International Conference on the Foundations of Digital Games* (pp. 1-8). ACM.

Bauckhage, C., Kersting, K., Sifa, R., Thurau, C., Drachen, A., & Canossa, A. (2012). How players lose interest in playing a game: An empirical study based on distributions of total playing times. In *IEEE Conference on Computational Intelligence and Games* (pp. 139-146). IEEE.

Belicza, A. (2014). *Scelight*. Retrieved from https://sites.google.com/site/scelight/

Blizzard Entertainment. (1998). *StarCraft: Brood War* [PC game]. Irvine, CA: Blizzard Entertainment

Bosc, G., Kaytoue, M., Raïssi, C., & Boulicaut, J. F. (2013). Strategic patterns discovery in RTS-games for e-sport with sequential pattern mining. In *MLSA@ PKDD/ECML* (pp. 11-20).

Bowman, B., Elmqvist, N., & Jankun-Kelly, T. J. (2012). Toward visualization for games: Theory, design space, and patterns. *IEEE Transactions on Visualization and Computer Graphics*, *18*(11), 1956-1968.

Bromley, S. (2018). Interviewing players. In A. Drachen, P. Mirza-Babaei, & Nacke, L. (Eds.), *Games user research* (pp. 163-174). Oxford, UK: Oxford University Press.

Brühlmann, F., & Mekler, E. (2018). Surveys in games user research. In A. Drachen, P. Mirza-Babaei, & Nacke, L. (Eds.), *Games user research* (pp. 141-162). Oxford, UK: Oxford University Press.

Bungie (2018). *The Bungie.Net API*. Retrieved from https://github.com/Bungie-net/api

Canossa, A., Nguyen, T. H. D., & El-Nasr, M. S. (2016). G-Player: Exploratory visual analytics for accessible knowledge discovery. In *Proceedings of the First International Joint Conference of DiGRA and FDG*.

Carr, H. (2015). *Game design: From data driven to data informed*. Presentation at the 2015 Game Developers Conference Europe. Retrieved from https://www.gdcvault.com/play/1022796/Game-Design-From-Data-Driven

Castronova, E., Williams, D., Shen, C., Ratan, R., Xiong, L., Huang, Y., & Keegan, B. (2009). As real as real? Macroeconomic behavior in a large-scale virtual world. *New Media & Society*, *11*(5), 685-707.

Chen, P. P., Guitart, A., del Río, A. F., & Periáñez, Á. (2018). Customer lifetime value in video games using deep learning and parametric models. Retrieved from https://arxiv.org/abs/1811.12799

Coenen, F. (2011). Data mining: past, present and future. *The Knowledge Engineering Review*, *26*(1), 25-29.

Cole, D. (2018). *Electronic Arts and Activision Blizzard focus on games as a service*. Retrieved from http://www.dfcint.com/dossier/electronic-arts-and-activision-blizzard-focus-on-games-as-a-service/

Dankoff, J. (2014). *Game telemetry with DNA tracking on Assassin's Creed*. Retrieved from https://www.gamasutra.com/blogs/JonathanDankoff/20140320/213624/Game_Telemetry_with_DNA_Tracking_on_Assassins_Creed.php

Desurvire, H., & El-Nasr, M. S. (2013). Methods for game user research: studying player behavior to enhance game design. *IEEE Computer Graphics and Applications*, *33*(4), 82-87.

Dong, G., & Pei, J. (2007). *Sequence data mining (Vol. 33)*. Springer Science & Business Media.

Drachen, A., & Canossa, A. (2009). Towards gameplay analysis via gameplay metrics. In *Proceedings of the 13th International MindTrek Conference* (pp. 202-209). ACM.

Drachen, A., El-Nasr, M. S., & Canossa, A. (2013). Game analytics–the basics. *In Game Analytics* (pp. 13-40). Springer, London.

Drachen, A., Thurau, C., Sifa, R., & Bauckhage, C. (2013). A Comparison of methods for player clustering via behavioral telemetry. In *Proceedings of the 8th International Conference on the Foundations of Digital Games* (pp. 245-252). ACM.

Drachen, A., Yancey, M., Maguire, J., Chu, D., Wang, I. Y., Mahlmann, T., ... & Klabajan, D. (2014). Skill-based differences in spatio-temporal team behaviour in Defence of the Ancients 2 (DotA 2). In *IEEE Games Media Entertainment* (pp. 1-8). IEEE.

El-Nasr, M. S., Desurvire, H., Aghabeigi, B., & Drachen, A. (2013). Game analytics for game user research, Part 1: A workshop review and case study. *IEEE Computer Graphics and Applications*, *33*(2), 6-11.

Fekete, J. D., van Wijk, J. J., Stasko, J. T., & North, C. (2008). The value of information visualization. In *Information Visualization* (pp. 1-18). Springer, Berlin, Heidelberg.

Fields, T., & Cotton, B. (2011). *Social game design: Monetization methods and mechanics*. CRC Press.

Fullerton, T., Swain, C., & Hoffman, S. (2004). *Game design workshop: Designing, prototyping, & playtesting games*. CRC Press.

Galyonkin, S. (2017). *Monthly summaries*. Retrieved from https://steamspy.com/year/

Hadiji, F., Sifa, R., Drachen, A., Thurau, C., Kersting, K., & Bauckhage, C. (2014). Predicting player churn in the wild. In *IEEE Conference on Computational Intelligence and Games* (pp. 1-8). IEEE.

Hamari, J., Alha, K., Järvelä, S., Kivikangas, J. M., Koivisto, J., & Paavilainen, J. (2017). Why do players buy in-game content? An empirical study on concrete purchase motivations. *Computers in Human Behavior*, *68*, 538-546.

Hanner, N., & Zarnekow, R. (2015). Purchasing behavior in free to play games: Concepts and empirical validation. In *48th Hawaii International Conference on System Sciences* (pp. 3326-3335). IEEE.

Hazzard, E. (2014). *Data visualization in games*. Retrieved from http://vasir.net/blog/game-development/data-visualization-in-games

Hoobler, N., Humphreys, G., & Agrawala, M. (2004). Visualizing competitive behaviors in multi-user virtual environments. In *Proceedings of the Conference on Visualization'04* (pp. 163-170). IEEE.

Hooshyar, D., Yousefi, M., & Lim, H. (2018). Data-driven approaches to game player modeling: A systematic literature review. *ACM Computing Surveys*, *50*(6), 90.

IGDA GRUX (2018). *What is GUR/UX?*. Retrieved from http://grux.org/what-is-gurux/

Kang, S. J., Kim, Y. B., Park, T., & Kim, C. H. (2013). Automatic player behavior analysis system using trajectory data in a massive multiplayer online game. *Multimedia Tools and Applications*, *66*(3), 383-404.

Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons.

Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization & Computer Graphics*, *8*(1), 1-8.

Keim, D. A., Mansmann, F., Schneidewind, J., & Ziegler, H. (2006). Challenges in visual data analysis. In *Tenth International Conference on Information Visualization* (pp. 9-16). IEEE.

Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In *Information Visualization* (pp. 154-175). Springer, Berlin, Heidelberg.

Kennerly, D. (2003). *Better game design through data mining*. Retrieved from https://www.gamasutra.com/view/feature/131225/better_game_design_through_data_.php

Kim, J. H., Gunn, D. V., Schuh, E., Phillips, B., Pagulayan, R. J., & Wixon, D. (2008). Tracking real-time user experience (TRUE): a comprehensive instrumentation solution for complex systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 443-452). ACM.

King, R., Churchill, E. F., & Tan, C. (2017). *Designing with data: Improving the user experience with A/B testing*. O'Reilly Media, Inc.

Koskenvoima A., Mäntymäki M. (2015) Why do small and medium-size freemium game developers use game analytics?. In Janssen M. et al. (Eds.), *Open and Big Data Management and Innovation* (pp. 326-337). Cham: Springer.

Kuan, Y. T., Wang, Y. S., & Chuang, J. H. (2017). Visualizing real-time strategy games: The example of StarCraft II. In *IEEE Conference on Visual Analytics Science and Technology*. IEEE.

Lehdonvirta, V. (2009). Virtual item sales as a revenue model: identifying attributes that drive purchase decisions. *Electronic Commerce Research*, *9*(1-2), 97-113.

Lynn, J. (2012). Data metrics and user experience testing. In *CHI 2012 Workshop on Game User Research*. Retrieved from http://gur.hcigames.com/wp-content/uploads/2015/02/Data-Metrics-and-User-Experience-Testing.pdf

Mansell, P. (2015). *Analytics in action: Data in a creative culture*. Presentation at the 2015 Game Developers Conference Europe. Retrieved from https://www.gdcvault.com/play/1022776/Analytics-in-Action-Data-in

Medler, B. (2009). Generations of game analytics, achievements and high scores. Eludamos. *Journal for Computer Game Culture*, *3*(2), 177-194.

Medler, B. (2011). Player dossiers: Analyzing gameplay data as a reward. *Game Studies*, *11*(1).

Medler, B., John, M., & Lane, J. (2011). Data cracker: developing a visual game analytic tool for analyzing online gameplay. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2365-2374). ACM.

Mirza-Babaei, P., Wallner, G., McAllister, G., & Nacke, L. E. (2014). Unified visualization of quantitative and qualitative playtesting data. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems* (pp. 1363-1368). ACM.

Mirza-Babaei, P., Moosajee, N., & Drenikow, B. (2016). Playtesting for indie studios. In *Proceedings of the 20th International Academic Mindtrek Conference* (pp. 366-374). ACM.

Missura, O., & Gärtner, T. (2009). Player modeling for intelligent difficulty adjustment. In *International Conference on Discovery Science* (pp. 197-211). Berlin, Heidelberg: Springer.

Morrison, M., & Fontenla, M. (2013). Price convergence in an online virtual world. *Empirical Economics*, *44*(3), 1053-1064.

Nacke, L. (2015). Games user research and physiological game evaluation. In R. Bernhaupt (Ed.), Game user experience evaluation (pp. 63-86). Cham, Switzerland: Springer.

Pedersen, C., Togelius, J., & Yannakakis, G. N. (2010). Modeling player experience for content creation. *IEEE Transactions on Computational Intelligence and AI in Games*, *2*(1), 54-67.

Polyphony Digital. (2010). *Gran Turismo 5* [Video game]. San Mateo, CA: Sony Computer Entertainment

Powell, R. (2016). *Positive and negative effects of game analytics in the game design process: A grounded theory study* (Master thesis, Uppsala University).

psychotrip (2016). *"The Ubisoft Formula": Repetition and stagnation in video games*. Retrieved from https://comicsverse.com/ubisoft-formula/

Remo, C. (2009). *Yamauchi: Gran Turismo 5's development cost hit $60 million*. Retrieved from http://www.gamasutra.com/view/news/116910/Yamauchi_Gran_Turismo_5s_Development_Cost_Hit_60_Million.php

Riot Games (2018). *Riot developer portal*. Retrieved from https://developer.riotgames.com/

Rockstar North. (2013). *Grand Theft Auto V* [Video game]. New York City, NY: Rockstar Games

Schoenblum, D. (2010). *Zero to millions: Building an XLSP for Gears of War 2*. Presentation at the 2010 Game Developer Conference. Retrieved from https://www.gdcvault.com/play/1012329/Zero-to-Millions-Building-an

Shneiderman, B. (2002). Inventing discovery tools: combining information visualization with data mining. *Information Visualization*, *1*(1), 5-12.

Sifa, R., Hadiji, F., Runge, J., Drachen, A., Kersting, K., & Bauckhage, C. (2015). Predicting purchase decisions in mobile free-to-play games. In *Proceedings of the Conference on Artificial Intelligence and Interactive Digital Entertainment* (pp. 79-85). AAAI.

Sinclair, B. (2013). *GTA V dev costs over $137 million, says analyst*. Retrieved from https://www.gamesindustry.biz/articles/2013-02-01-gta-v-dev-costs-over-USD137-million-says-analyst

Sony Online Entertainment. (1999). *EverQuest* [PC game]. San Diego, CA: Sony Online Entertainment.

Stahlke, S. N., & Mirza-Babaei, P. (2018). Usertesting without the user: Opportunities and challenges of an AI-Driven approach in games user research. *Computers in Entertainment*, *16*(2), 9.

Synnaeve, G., & Bessiere, P. (2011). A Bayesian model for opening prediction in RTS games with application to StarCraft. In *IEEE Conference on Computational Intelligence and Games* (pp. 281-288). IEEE.

Takahashi, D. (2017). *Data-driven versus intuition-driven game design*. Retrieved from https://venturebeat.com/2017/02/19/data-driven-versus-intuition-driven-game-design/

Temmerman, J. (2017). *World of Tanks replay parser*. Retrieved from https://github.com/evido/wotreplay-parser

Thompson, C. (2007). *Halo 3: How Microsoft labs invented a new science of play*. Retrieved from https://www.wired.com/2007/08/ff-halo-2/

Thomson, R., Lebiere, C., & Bennati, S. (2014). Human, model and machine: a complementary approach to big data. In *Proceedings of the 2014 Workshop on Human Centered Big Data Research* (p. 27). ACM.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, PA: Addison-Wesley.

Wallner, G., & Kriglstein, S. (2012). A spatiotemporal visualization approach for the analysis of gameplay data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1115-1124). ACM.

Wallner, G. (2013). Play-Graph: A methodology and visualization approach for the analysis of gameplay data. *In Proceedings of the 8th International Conference on the Foundations of Digital Games* (pp. 253-260).

Wallner, G., & Kriglstein, S. (2013). Visualization-based analysis of gameplay data–a review of literature. *Entertainment Computing*, *4*(3), 143-155.

Wallner, G., & Kriglstein, S. (2014). PLATO: A visual analytics system for gameplay data. *Computers & Graphics*, *38*, 341-356.

Wallner, G. (2015). Sequential analysis of player behavior. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play* (pp. 349-358). ACM.

Wallner, G., & Kriglstein, S. (2015). An introduction to gameplay data visualization. In P. Lankoski & S. Björk (Eds.), *Game research methods* (pp. 231-250). ETC Press.

Wallner, G. (2018). Automatic generation of battle maps from replay data. *Information Visualization*, *17*(3), 239–256.

Wargaming. (2010). *World of Tanks* [PC game]. Minsk, Belarus: Wargaming

Weber, B. G., & Mateas, M. (2009). A data mining approach to strategy prediction. In *IEEE Symposium on Computational Intelligence and Games* (pp. 140-147). IEEE.

Weber, B. (2018). *A history of game analytics platforms*. Retrieved from https://www.gamasutra.com/blogs/BenWeber/20180409/316273/A_History_of_Game_Analytics_Platforms.php

Whitson, J. R. (2012). *Game design by numbers: Instrumental play and the quantitative shift in the digital game industry* (Doctoral dissertation, Carleton University).

Williams, D. (2015a). *User acquisition and attribution: The basics (Part 1)*. Retrieved from https://www.gamasutra.com/blogs/DmitriWilliams/20150304/237833/User_Acquisition_and_Attribution_The_Basics_Part_1.php

Williams, D. (2015b). *The tenets of community management series: Part 1 of 3*. Retrieved from https://www.gamasutra.com/blogs/DmitriWilliams/20150528/244497/The_Tenets_of_Community_Management_Series_Part_1_of_3.php

Xie, H., Devlin, S., Kudenko, D., & Cowling, P. (2015). Predicting player disengagement and first purchase with event-frequency based data representation. In *IEEE Conference on Computational Intelligence and Games* (pp. 230-237). IEEE.

Zimmerman, E. (2015). Manifesto for a ludic century. In S. P. Walz & S. Deterding (Eds.), *The gameful world: Approaches, issues, applications* (pp. 19-22). Cambridge, MA: MIT Press.

Zoeller, G. (2010). *Development telemetry in video games projects*. Presentation at the 2010 Game Developer Conference. Retrieved from https://www.gdcvault.com/play/1012227/Development-Telemetry-in-Video-Games

Zook, A., & Riedl, M. O. (2012). A temporal data-driven player model for dynamic difficulty adjustment. In *Proceedings of the Conference on Artificial Intelligence and Interactive Digital Entertainment* (pp. 93-98). AAAI.