UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE EFFECT OF GATHERING ON SANDBOX PLAYER ENGAGEMENT AS DEFINED USING ANALYTIC METHODS

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

EMILY KATHRYN GRIMES
Norman, Oklahoma
2017

THE EFFECT OF GATHERING ON SANDBOX PLAYER ENGAGEMENT AS DEFINED USING ANALYTIC METHODS

A THESIS APPROVED FOR THE
GALLOGLY COLLEGE OF ENGINEERING

BY

____
Dr. Charles Nicholson, Chair

____
Dr. Randa Shehab

____
Dr. Ziho Kang

Table of Contents

List of Tables

List of Figures

Abstract

Player engagement is a concept that is both vital to the online gaming industry and difficult to define. Typically, engagement is defined using social science methodology such that observing, surveying, and interviewing players are commonly implemented. Further, as online gaming increases in popularity, social behavior in games is also increasingly prevalent. This phenomenon is also studied most often by social scientists. With the vast amount of data being collected from video games as well as user bases increasing in size, it is worthwhile to investigate whether or not user engagement can be defined and interpolated from data alone. This study develops a methodology for defining engagement using analytic methods in order to approach the question of whether *gathering* in sandbox games has an effect on player engagement.

Chapter 1: Introduction

Online gaming is more prevalent with half of American adults reporting that they play video games (Duggan, 2015). In the United States, the video game industry generated 30.4 billion dollars in revenue in 2016; worldwide, the industry made up to 99.6 billion dollars (ESA, 2017; Newzoo, 2016). Online games especially depend on engaged users to make money (Eastin, Daugherty, & Burns, 2010). One popular tactic for online game developers is to make games addictive enough that users incorporate them into their normal routines, returning to the game every single day (Needleman, 2017). Indeed, engaged users are those who invest their time, energy, and emotions into a product (Attfield et al., 2011). User engagement has been further defined as the "emotional, cognitive, and behavioral connection that exists, at any point in time and possibly over time, between a user and a resource" (Attfield et al., 2011). Engagement is a quality of user experience that can be quantified, but to measure it properly, engagement must be broken down to a list of its characteristics that are quantifiable. Some indicators of engagement include focused attention, positive emotions felt by users, visual and sensory appeal, likelihood users will remember an experience and want to repeat or recommend it to other potential users, novelty, resource reputation, user motivation, incentives, and benefits, challenge, interactivity, and feedback (Attfield et al., 2011; O'Brien & Toms, 2008). From these it is clear that engagement cannot be entirely quantified in an objective manner, but features may be generated from existing data that can represent indicators of engagement. These features may then be incorporated into a metric for engagement that approximate the state of engagement.

Massively multiplayer online games (MMOGs) have been referred to as "petri dishes for social science" (Castronova, 2006). As will be described in section 2, models for the state of being engaged have been iterated over by social scientists studying human users of digital environments. Likewise, user interactions in the digital world have been of particular interest to scholars of the humanities. Many dynamics of social interaction have been studied in the context of *World of Warcraft (WoW),* for example. *WoW* was one of the earliest role-playing MMOGs that remains popular today. Collaboration, personality expression, and even racism are among the topics scholars have addressed in the context of social behavior in *WoW* (Monson, 2012; Nardi & Harris, 2006; Yee, Ducheneaut, Nelson, & Likarish, 2011). Research of this kind is valuable because it helps the human race understand itself better as lifestyles become increasingly dependent on digital resources, particularly as those resources become ever more immersive experiences. On a more pragmatic note, the prevalence of this kind of research demonstrates that social interaction in MMOGs is a widespread phenomenon. Quantitative approaches to understanding social activity in MMOGs are less common in the literature, but are still potentially profoundly insightful. This work is less interested in the quality of interactions, but rather in the potential for those interactions and the effect that may have on what we come to define as engagement.

This study observes a case study of player behavior within the beta release of an open-world sandbox game soon to be released by Nerd Kingdom known as *The Untitled Game* (*TUG*). We ask, is there a quantifiable relationship between the opportunity for social aspects of player behavior and engagement? It is likely that the answer to this question may differ from player to player, and it may differ from game to game as well.

Observations from *TUG* may illuminate patterns present in similar games with respect to the fact that players are able to sign on to the same servers simultaneously such as *Minecraft* and *WoW*. The case study presented here provides methodological guidance on defining terms like engagement from an analytics perspective. The paper is organized as follows: Section 2 is a brief overview of prior research, Section 3 provides a methodological framework for the case study, Section 4 covers the methods and results of the case study, and Section 5 is a concluding discussion of insights from the case study results.

Chapter 2: Literature Review

User engagement research is essential to the success of all online web applications. Research in user engagement for online games has been especially prolific in recent years. As metrics and methodologies for defining and validating the state of engagement for users propagate, it is increasingly necessary to investigate factors that impact that state. As an increasing number of individuals immerse themselves in virtual worlds, understanding how human interaction within the gaming environment affects user engagement is also an important area of research from a commercial perspective, if not also from an anthropological perspective. In this section is a brief review of prior research on engagement in MMOGs in areas of defining engagement in the context of these games, observing the phenomenon of player-on-player interaction within the games, and finally, game analytics approaches to gaining insight about player behavior in MMOGs.

Most academic discussion of user engagement since the 1990s begins with work that models the state of being engaged using methodologies deeply rooted in social science, such as surveys and field observation. For example, O'Brien & Toms (2008) built on this body of work to develop a detailed model of engagement as a process in time beginning with a point of engagement and ending on a moment of disengagement over sensual, emotional, and spatiotemporal threads of experience. In 2010, O'Brien & Toms (2010) developed a survey instrument comprised of six factors: Perceived Usability, Aesthetics, Novelty, Felt Involvement, Focused Attention, and Endurability which they called the User Engagement Scale (UES) for online shopping contexts. Wiebe et al. (2014) extended this work ultimately revising the UES to the context of games. Factors analyzed to create this modified UES, coined UESz, were: Focused Attention, Perceived Usability on the part of the user, Aesthetic Affect, and Overall User Satisfaction. Data for this study was collected via Amazon's Mechanical Turk. Users played a game for a minimum amount of time and took a survey afterwards. In fact, surveys are the predominant way that player engagement is measured. In 2009, the Game Engagement Questionnaire was formally developed (Brockmyer et al., 2009), and by 2016 the Game User Research (GUR) community had enough questionnaires that it was deemed appropriate to move towards a single questionnaire that aggregated the best qualities in the most popular questionnaires—which still included the Game Engagement Questionnaire (Denisova, Nordin, & Cairns, 2016). Alternative methods to player surveys for investigating player engagement are less common, but Kirschner & Williams (2015) developed the Gameplay Review Method (GRM). GRM goes beyond surveys; instead, GRM relies on in-depth interviews of players as well as analysis of audiovisual recordings of gameplay. The GRM links empirical and interpretive data to inform game design with a comprehensive, if not holistic, understanding of player engagement.

It is difficult to prove whether users are engaged or not with a resource, but fortunately many have succeeded in doing so. For example, Lehmann et al. (2012) constructs a model to measure user engagement on the web with online behavior metrics. Schoenau-Fog & Henrik (2014) explores the player engagement process in games by identifying components associated with players' desire to keep playing.

The game analytics community is a growing part of the GUR community at large. Though analytics approaches to the question of engagement as less common in the literature compared to other approaches mentioned above, analytics approaches are gaining popularity. An analytics approach does not require researchers to get to know players or assess their subjective responses to questionnaires. Analytics methods are rooted in business intelligence practices instead. The challenge with these methods, then, is that they rely on data logged from user behavior in a resource. Since the concept of engagement is somewhat difficult to quantify and users are typically anonymous or unavailable for qualitative follow-up, definitions for engagement or similar outcomes must be determined from whatever data is present. Kawale et al. (2009) are among those who have tried to assess how engaged users are by predicting when they will leave the game, or "churn". While it may be difficult to quantify the quality of being engaged, it is relatively straight-forward to determine when a player is no longer active. With this in mind, analysts can apply predictive modeling to user log data from game play events and try to predict phenomena like churn.

Kawale et al. (2009) construct a graph of players of *EverQuest II* (*EQ2*) where an edge exists between two players if they participated in a quest together. Edge weight is determined by the number of points the players shared. Muller et al. (2015) also use graphs to quantify and predict collaboration in *Minecraft*. While collaboration is not engagement, it is a concept that requires a clear definition to facilitate empirical analysis. To quantify collaboration, they construct undirected, weighted graphs where the vertices are players and the edges represent one of several collaboration indicators they identify including contact and chat. These indicators are defined further. For example, two players are considered to be in contact if both are active and the distance between them is 15 blocks or less apart. Players are said to have chatted if conversations between players are detected in the native messaging area of *Minecraft*. The edges of a graph composed of these attributes are summed, multiplied by their weights, and divided by the duration of active play time to generate a single collaboration index for players in this study. Raimbault et al. (2016) defined three engagement levels on a session basis by applying *k*-means clustering to both session lengths and number of events per session in their data set. Beyond the above examples, overall playtime is often considered the best proxy for engagement in analytics contexts (Drachen, Thurau, & Bauckhage, 2013).

It would be an oversight to omit user motivation from a discussion about the user state of being engaged. Yee (2006) created a popular model of player motivations based off of replies from surveys filled out by players of a wide variety of games. Among the principal components of this model are the overarching motivations that were named Achievement, Immersion, and Social. Within these, there are more specific motivations. A socially motivated player, for example, may be more interested in actual socializing within the game, building relationships in particular, or of being part of a team more broadly. It is important to note that these motivation components are neither mutually exclusive nor do they suppress other motivation components. This social aspects of games can be a user's sole motivation to engage with a game, or these aspects may just be part of a player's motivation profile. Thus, it is understood that social aspects of games contribute to player motivation, a documented indicator of engagement.

Social aspects of player behavior have been studied in GUR. Kawale et al. (2009) determine that as a player's "neighbor"—a term they define in more detail—

churns, her probability of churning increases. According to Kawale et al., churn behavior in *EQ2* has a social component. Likewise, Zhuang et al. (2007) performed an in-depth study of player dynamics in *WoW*. In their work, they determine three predictors of longer session lengths to be player level, start time of the session, and having played long sessions previously. Players in their research who achieved higher levels were more likely to play longer. Furthermore, sessions started in the evening were more likely to last longer than sessions started in the morning, between 5 and 7 A.M. in particular. Session length is a commonly used indicator of player engagement. Understanding factors that influence session length can aide analysts in defining engagement metrics for their particular game's dataset as well as design data structures that will allow for more accurate definition of engagement. Beyond session length, Zhuang et al. observed interactivity between players, which they claimed had little effect on other session attributes. Their study was based on 1000 players observed for 5 months with collected data attributes including session length, downtime, inter-arrival times, availability, aggregate churn rate, and degree of player independence. Pirker et al. (2016) analyzed the effect of social networks in the hybrid online shooter game, *Destiny*, on player behavior and found that players with a tendency to play with the same people play better with respect to win/loss and kill/death ratios. These results indicate that social interaction in the game *Destiny* may have a positive impact on player engagement since increased performance may have a positive impact on factors of engagement such as player mood. Finally, Ducheneaut et al. (2006) took an early look at social dynamics in MMOGs using *WoW* data and determined that instead of forming relationships, players tended to use other players as audience for their performances. Ducheneaut et al. (2006) find that designing for a "spectator experience" may be more valuable for game designers than to assume players are socially motivated.

Game analytics perspectives on quantifying terms like engagement that are possibly more in the wheelhouse of social scientists are discussed. Also worth discussion are the user behavior analyses performed by academic game analysts. Of these studies, social aspects of user behavior are underrepresented in the literature. Raimbault et al. (2016) applied clustering techniques to user session logs from *TUG*. Clustering was applied to session logs with only a few attributes about players' actions. Sessions were broken into groups defined by the characterizing behavior of the players for that particular session. The work of Drachen et al. (2012) precedes that of Raimbault et al. in that clustering of telemetry data from MMOG players was the focus, but instead of one game, significantly more massive player data sets from two games were analyzed. Overall, the paper is an introduction to classifying player data. Likewise, Drachen et al. (2013) provides an explicit overview of clustering methods that work well with player telemetry data. While Drachen's work predominates the past decade, early work on player behavior analytics dates at least to 2005. Kwok & Yeung (2005) examines player behavior in an early MMOG called *RockyMud*. The primary motivation for this analysis was not player experience or satisfaction but system performance. Thus, attributes analyzed were related to time spent on a server and location in the game. Even so, the analysis of player location in this work leads to another recent avenue of GUR from the game analytics perspective—spatial analysis. The intent of spatial analysis is to not only understand what the players are doing, but to be able to visualize this behavior in the context of the virtual environment. In Drachen

& Canossa (2011), spatial user behavior is the subject in terms of player motion, but the work also serves mostly as an overview and a call to action for research. Very little work with spatiotemporal analysis of user data from games has been published to date. As Drachen points out, there are proprietary tools built in-house at large game companies for this kind of analysis. These tools and the work that follows from them benefit only the developers for the games made by those companies and not outside researchers in the field or academia. Even so, spatial analysis is a promising area for game analysts to explore player interactions via telemetry data in lieu of or in addition to more traditional social science methodologies utilized in studies mentioned above. See Drachen & Schubert (2013) for a summary of the work to-date in the field of spatial analysis for GUR.

The present study is positioned among the work of Drachen & Schubert, Kawale et al., Muller et al., Raimbault et al., and Zhuang et al. (2013; 2009; 2015; 2016; 2007). We observe a case study of player behavior within the beta release of *TUG,* also analyzed by Raimbault et al. We define engagement using an index derived with methods based loosely on those used to define collaboration by Muller et al., and we use spatial analysis to determine a proxy for user interaction. We ask, is there a quantifiable relationship between the opportunity for social aspects of player behavior and engagement? Zhuang et al. (2007) claims interactivity between players has little effect on other session attributes in the *WoW*, yet Kawale et al. (2009) takes the nearly opposite position that game churn behavior is associated with the social component of a game. Churn and engagement are not the same, but churn sometimes serves as a proxy for engagement, or rather an indicator that engagement has ended. It is likely that the answer to this question may differ from player to player, and it may differ from game to game as well. Even so, the case study presented here will contribute to the on-going conversation by providing insight on whether players who had the opportunity to interact in the beta release of *TUG* were less likely to have churned. Further, this work provides methodological guidance on defining terms like engagement from an analytics perspective.

Chapter 3: Methodology

Using Telemetric Data to Define Terms

Engagement is one of the most popular metrics used to rationalize monetization for new technologies, especially web applications and video games (Eastin et al., 2010). It is not surprising that engagement is a common subject in GUR. Even so, GUR studies focusing on engagement are typically rooted in social science methodology. Researchers do not always have access to players for interviews, surveys, or observation, especially when those researchers are data analysts working with data collected from user actions taken in the resource. Likewise, social interaction in games is also frequently approached from a social science perspective. Researchers will often get data by becoming a player and observing others at play. But again, for data analysts, this methodology isn't necessarily practical. Thus, a methodology for defining terms with qualitative connotations grounded solely in analysis of telemetric data is called for.

Defining Engagement

It is a challenge to frame the concept of engagement in quantitative terms due to the fact that the term engagement includes subjective attributes that are difficult to measure. User logs for most digital resources do not include any fields for motivation, appeal or affect, or even positive emotions, for example. Instead, studies that have analyzed player engagement solely from user logs tend to make the assumption that playtime is the best proxy for engagement (Drachen et al., 2013). But it may possible to do better. As mentioned above, engagement is one of the most popular metrics used to rationalize monetization; therefore, the best analysis of engagement in games should utilize the research on the concept and try to define engagement in ways that incorporate as many indicators of this quality as is rational.

A framework for assessing player engagement in contexts where data is provided to an analyst (rather than collected in the course of a designed experiment) that goes beyond the use of playtime to assess this state should prioritize capturing as many indicators of the concept of engagement as possible without redundancy. Thus, a feature for engagement generated from other indicators will have two main qualities: (i) A high correlation with as many features corresponding to indicators of engagement as are available, and (ii) the features it correlates highly with will not be highly correlated with each other if that is avoidable. For example, there may be a high correlation between the number of clicks in a game and the amount of time a player spent playing, but these are both potentially valuable indicators of engagement that should be represented by an engagement metric.

Defining Interaction

It would be ideal for games where user interaction is possible—that is players are able to communicate as well as be in the same place, at the same time, on the same server, and see each other—that data logged from them will include at least one feature to indicate that players interacted by design. If not, there may be ways to determine that players had the opportunity to interact from spatial analysis. If chat logs can be joined with behavior logs on a timestamp, or similar attribute, then for some data sets determining player interaction in terms of "did interact" or "did not interact" can be straightforward. That said, spatial analysis of user coordinates in 3D games is an excellent way to visualize and assess if two players were in the same place at the same time. Geographic information systems (GIS) like ESRI ArcGIS and the open-source QGIS are useful for this purpose. Drachen (2013) points out the usefulness of GIS for behavioral analysis. By determining that users were in the same proximity, even if there is no way to prove that they knew the other was present at the time and place they were, one can assert that the opportunity for social interaction was present.

Chapter 4: Case Study

In 2013, Nerd Kingdom, a software company based in Texas, launched a *Kickstarter.com* campaign to raise capital for *The Untitled Game*, or *TUG*. Funds were successfully raised, and development on *TUG* has continued ever since. Plans for an official launch of the game are set for 2017. *TUG* is a sandbox—a style of game in which the narrative is non-linear. The gamer is allowed to roam freely and interact with the virtual world at will. While most sandboxes encourage exploration with little or no forced game progression, one of *TUG's* distinguishing characteristics will be that player experience will be heavily dependent on outcomes produced by artificial intelligence (AI). An AI engine will learn from a player's behavior and change his experience to optimize engagement for that player. This case study is based on a limited data set collected in 2016 during a short-term beta test of *TUG*. To be invited to participate in this beta test, users had to spend at least 30 dollars to support the creation of *TUG* on *Kickstarter* (Nerd Kingdom, 2013).

Data from the beta test represent events logged for game sessions from a 42-day period, January 6, 2016 to February 17, 2016, involving 89 users—only 82 of which were associated with events beyond logging in and out. A session is defined by a series of events that occur over a period of time between log-in and log-out or idle events. There are 315,307 rows of data covering 553 unique sessions where each row represents a single event occurring within a session. Not counting log in or log out events, there are 314,483 rows of data. Attributes per event are a session ID number, user name, a date-time stamp, and game coordinates for x, y, and z planes. Coordinates represent player location in *TUG*. Indeed, these coordinates do not have meaning outside the context of *TUG* itself but can be used to create a *TUG* geography, or a map of the *TUG* world, by drawing a convex hull around the extreme coordinates plotted on a 2D plane. Since each row represents one event, each point represents a row of data and, likewise, one game event.

Table 1 Data from the beta test represent events logged for game sessions from a 42-day period, January 6, 2016 to February 17, 2016, involving 89 users. A session is defined by a series of events that occur over a period of time between log-in and log-out or idle events. Not counting log in or log out events, there are 314,483 rows of data.

Collected Data	
Attribute	
Events	315307
Sessions	553
Days	42
Users	89
Active Users	81
Sessions with Action	464
Action Events	314483

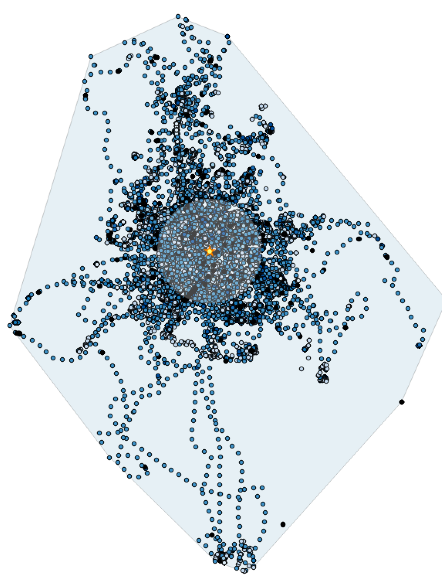

Figure 1 A geography of the *TUG* world can be created by drawing a convex hull around the extreme coordinates plotted on a 2D plane. Since each row represents one event, each point represents a row of data and, likewise, one game event. The yellow star represents the mean center of all the events, a point that likely approximates the default spawn location for the beta version of *TUG*.

Defining Engagement

There is not just one way to define engagement in the context of this data. It is worthwhile to explore multiple definitions to see which are more robust. In this case, we would call a definition more robust with respect to other definitions if it is highly correlated with more features corresponding to indicators of engagement and if those features are not redundant to each other. For this data in particular, one definition of engagement has already been developed. Raimbault et al. (2016) used *k*-means clustering to break sessions into three engagement categories which they describe as low, high, and very high engagement categories. The three clusters were formed on the lengths of sessions and the number of events per session. For the sake of clarity, we will refer to this definition as Definition 1 (D1). D1 allows for an objective partition to be created with respect to how relatively active a session is and how long a session lasts. Intuitively, D1 implies that busier and longer sessions will be considered more engaged than shorter and less active sessions. However, in the context of sessions, the correlation between the number of events per session and the length of a session in *TUG* is weak in the broad context of the data at 0.11. Removing outliers on session length strengthens the correlation between activity and session length to 0.6, but removing outliers on number of events per session makes the correlation lower than that of the original data at 0.09. It is notable that removing outliers on number of events per session is equivalent to removing all outliers on both attributes. The true relationship between session length and number of events per session is not as strong as D1 suggests; the three groups determined by D1 seem to be biased in favor of the length of the session. Running the *k*-means algorithm on data with respect to both session length and number of events per session creates groups that are not clearly distinct. In general, sessions that last longer do tend to have more events, but again, since the relationship is weak in the context of sessions, there are many sessions that are shorter but more active, or very long and less active. The longest sessions are given the highest engagement category, but some of the most active sessions are not clustered as highest engagement. Running the *k*-means algorithm on the same data with respect to session length alone, we see an almost identical clustering. Further, running *k*-means with respect to number of events per session alone produces clusters that are more clearly defined and intuitive in that the most active sessions are highly engaged, and long sessions with low activity are low engagement. D1 requires a trade-off between two of the few quantifiable elements of engagement we can pull out of the data, which means to use it, we end up favoring one of these elements (session length) instead of encapsulating both session length and number of events per session. It is worth considering this compromise since it may be that more eventful sessions contribute to profits as much if not more than sessions that last longer. A robust definition of engagement will sacrifice as little information that is profitable in the long run as is possible.

Table 2 The correlation between the number of events per session and the length of a session in *TUG* is weak in the broad context of the data. Removing outliers on session length strengthens the correlation between activity and session length, but removing outliers on number of events per session makes the correlation lower than that of the original data. It is notable that removing outliers on number of events per session is equivalent to removing all outliers on both attributes.

Correlation Between Number of Events Per Session and Session Length		
Data:	**Correlation:**	**Portion of Data Removed:**
no outliers removed	0.11	0
all outliers removed	0.09	0.08
removing outliers on session length only	0.6	0.01

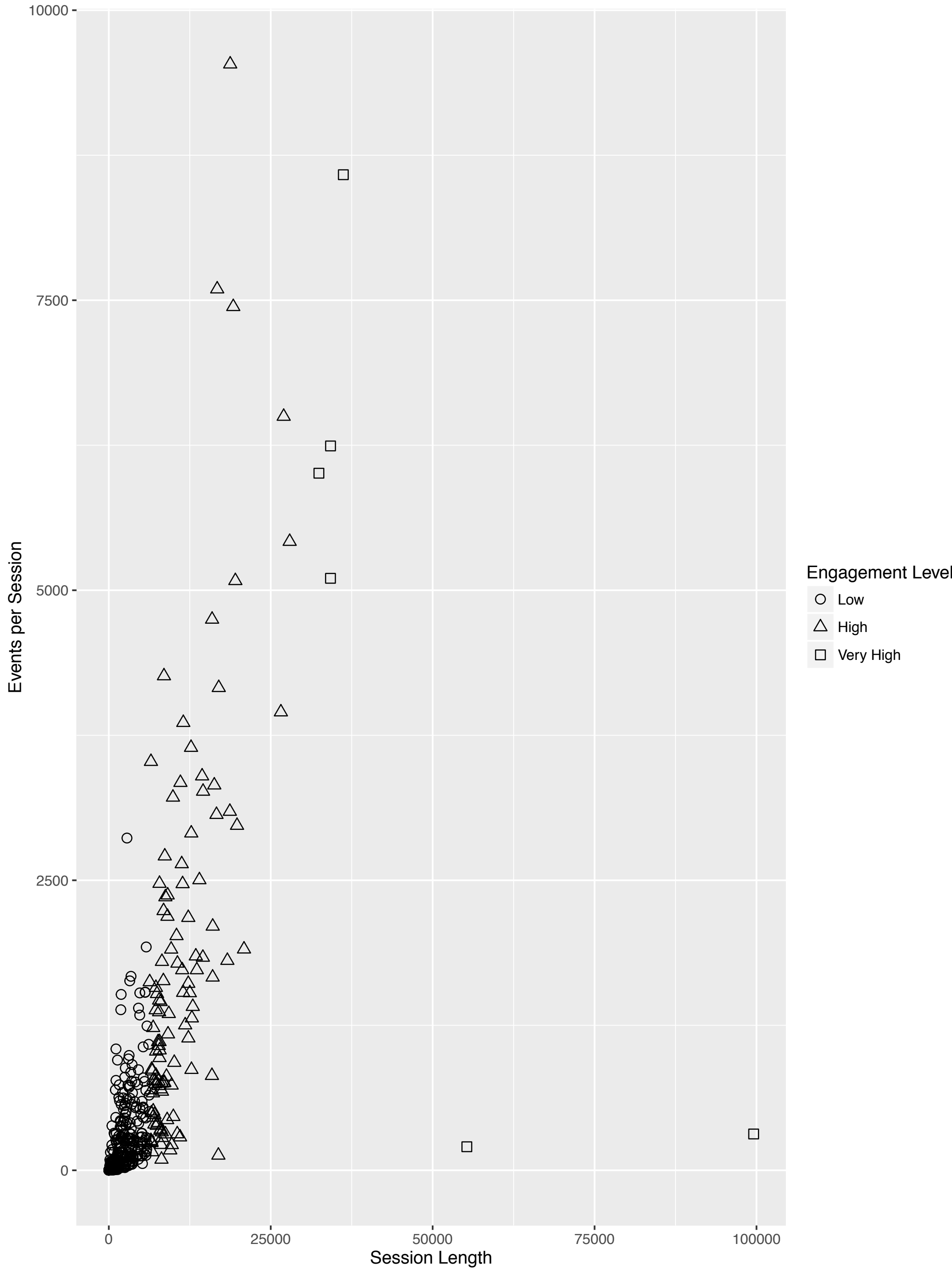

Figure 2 The three groups determined by D1 seem to be biased in favor of the length of the session. The above clusters were created with respect to both session length and events per session. In general, sessions that last longer do tend to have more events, but there are many sessions that are shorter but more active, or very long and less active. The longest sessions are given the highest engagement category, but some of the most active sessions are not clustered as highest engagement.

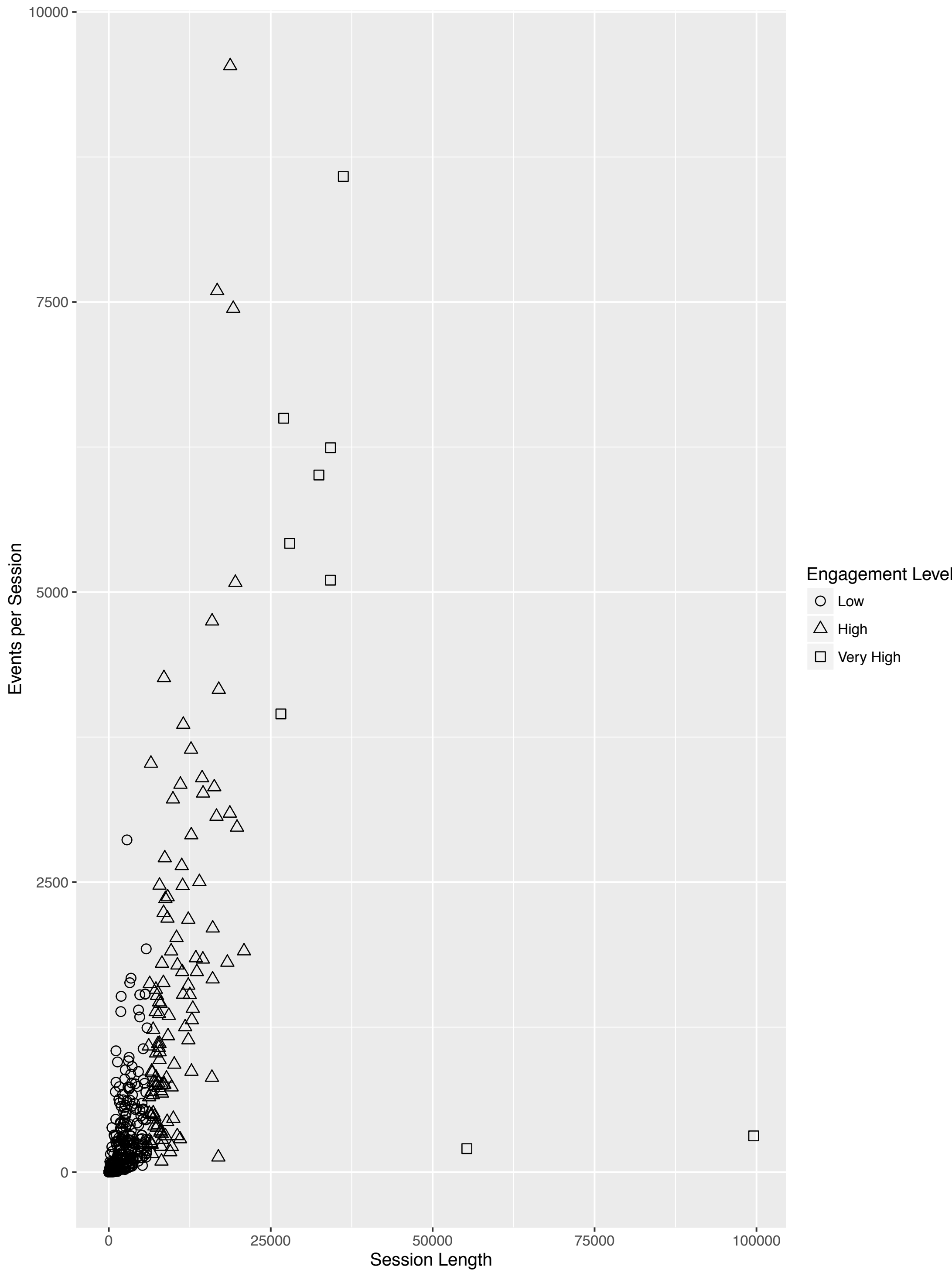

Figure 3 The three groups determined by D1 seem to be biased in favor of the length of the session. The above clusters were created with respect session length only. Differences between these clusters and those taken on both session length and events per session are negligible.

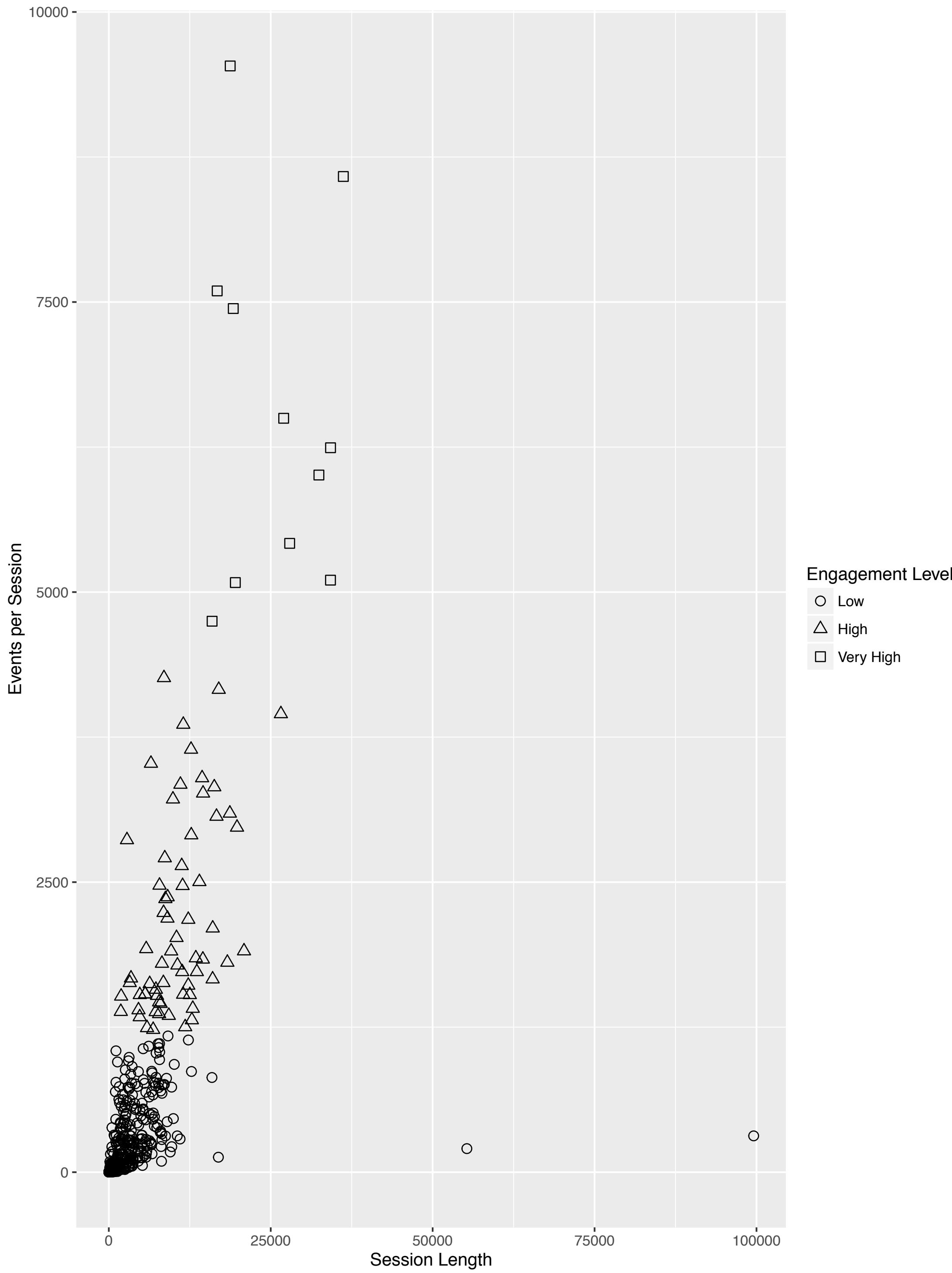

Figure 4 The three groups determined by D1 seem to be biased in favor of the length of the session. The above clusters were created with respect to events per session only. These clusters that are more clearly defined and intuitive in that the most active sessions are highly engaged, and long sessions with low activity are low engagement.

While D1 may not be as robust as is ideal, other definitions are not immediately intuitive. Regardless, our driving question—how the opportunity for social aspects of player behavior affects engagement—is user-based, and D1 represents a model of engagement that is session-based. To adapt to a user-centric model, it makes sense to explore the broad behavior of users instead of session-level behavior.

At the user level, it is clear that most users did not spend a lot of time in the game, regardless of activity levels, so the framework for a user-based model of engagement in this case has to begin by considering which users returned to the game. Half of users accumulated an hour or more of playtime, and about a quarter of users returned for a third session. About 30% of users played the game on more than two distinct days. That is, most users sat down to play one time, and the amount of time they spent during that one setting was their total accumulated playtime. A few users returned after logging out or going idle, which means they chose to continue playing after leaving. Likewise, users who played on more than two distinct days also chose to return to the game after going idle or logging out. We specify that users returned on more than two distinct days because some users logged on for the first time late enough on their first day that their single session ran into a second day despite overall playtime being relatively short. Thus, users who played on more than two distinct days are users who left and returned to play on a different day than the first session they ever played. With this in mind, Definition 2 (D2) was developed as a specific intersection of users among those possessing a total play time above a certain threshold, a relatively high number of active days, and a relatively high number of sessions for which the player returned. If a player's total play time was among the top half of total play times, and the player was in the top quartile for number of sessions played, and the player was active for more than two days, she is considered engaged in this model. There were 41 players (half of the active users) who accumulated more than one hour of play time, 21 players who logged in for five or more sessions, and 26 players who returned to play on three days or more. There were exactly 16 players who were among all of these groups. Again, the intent of D2 is to place emphasis on the likelihood that a user is going to return to the game.

Like D1, D2 is biased with respect to time. D2 is biased towards users who played on a greater number of unique days and who played more sessions. Even so, if engaged, users by this definition were more likely to return if they went idle or logged out, where less engaged users were more likely to ultimately churn upon idleness or logging out. D1 fails to accurately reflect that the most active sessions are the most engaged. Likewise in outcome, D2 also does not reflect that users with busier sessions are more engaged. That said, D2 does not claim to do so either, so its primary advantage on that point with respect to D1 is its transparency. Regardless, the absence of this information is just as much of a trade-off for D2 as it is for D1.

Another distinction of D2 from D1 is that it does not break down engagement into levels of low, high, or very high engagement. Instead, D2 identifies that a user is engaged or not in binary outcome. A weakness of the intersectional definition is that it may be somewhat redundant. While it appears to capture more indicators of engagement per user, there is a strong correlation (0.87) between the number of sessions a user played and the total unique days a user was active. It is obvious that D2 is no more ideal than a session-based model that cannot easily be applied to a user-based question.

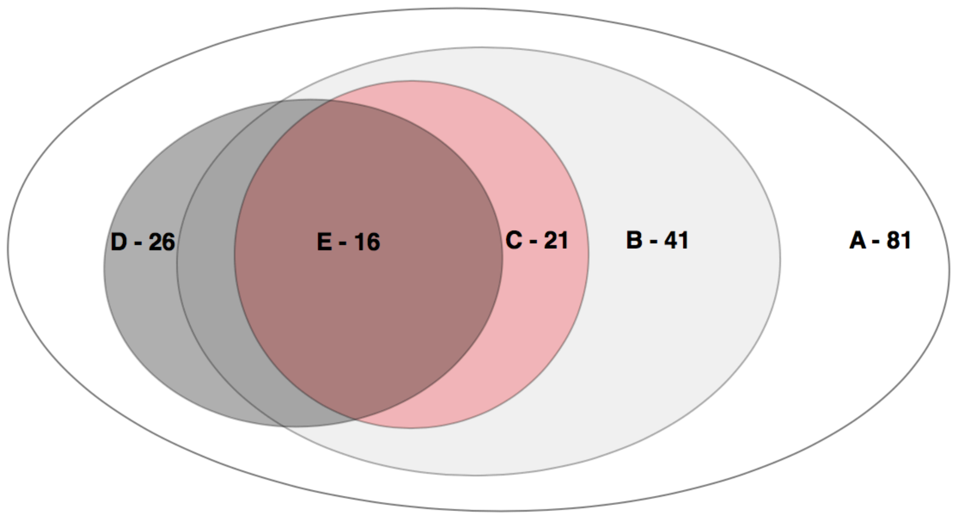

Figure 5 A: All active players. B: 41 (half of all) acive players accumulated more than an hour of playtime. C: 21 players logged in for 5 or more sessions. D: 26 players logged in on greater than 2 distinct days. E: The group of 16 players in the intersection of B, C, and D are said to be "engaged" by Definition 2.

Some authors, like Kawale et al. (2009), have relied on churn, the moment a player leaves the game, to inform their analyses instead of trying to define engagement. Building off of this, we now introduce a new engagement index (EI). EI is defined as the sum of the inverse probabilities of churn (Equation 1), or what we call the "percent engaged", of all events performed by a user (Equation 2). For this case study, we have the benefit of hindsight because our data was collected after the case study ended. We know exactly when each player churned for each session and when they ultimately churned. Therefore, the simplest way to calculate the probability of churn per session in our case is to determine how long each event per session is from the moment of churn.

Therefore, the inverse of this probability is the distance in time of an event from the moment of the session churn in proportion to the total amount of time the session lasted subtracted from 1. So, the first event of a session is 100% engaged in this scenario, because the probability of churn at the moment of log in is given as 0, and the last event in a session is 0% engaged, because the probability of churn at the moment of log out is given as 1. Engagement decreases over time until the moment of churn. This is equivalent to assigning log in events a value of 1 and weighting log out events a weight of 0. Future implementations of EI would weight other event types as well, given associated probabilities of churn determined by closer analysis. In this instance of EI, its value for each user accumulates to higher indices for users who were more active for longer, or who were more active over a longer period of time.

(1) $$p(churn) = 1 - \frac{D(e_n) - D(e_i)}{T(s)} \quad \forall s \in S, \forall e \in E,$$

where S is sessions, E is events, D is the date-time of each event, and T is the total time for the session.

(2) $$EI = \sum_{i=1}^{n}\left(1 - p(churn_e)\right),$$

where $p(churn_e)$ is the probability of churn for a specific event.

EI encapsulates three intuitive factors of engagement that we can easily quantify: session length, activity levels of sessions, and likelihood that users returned to the game. It is a user-based metric reflecting session-level information. D1, a purely session-based definition, does not map to the user level since most users had at least one engaged session. To translate this definition to the user level a metric similar to EI has to be made. One advantage of a numeric index such as EI is that clustering can be applied to create objective groupings of relative engagement. One of the shortcomings of clustering on EI in this case is that players' EI values do not form obvious groupings beyond a certain threshold. A majority of users have EI values of 200 or less, about a third of users have a range of EI values from 200 to just under 10,000, and exactly four players have EI of 10,000 or more. With differences so vast, the most intuitive clusters could be seen by suggesting that a *k*-means algorithm look for six centroids, and merging the resulting groups by eye. Players with EI less than 100 would certainly be low engagement, but less clear is how to identify engagement levels of EI between in the middle range. Using three centroids alone, *k*-means produces clusters in which about 90% of players are low engagement, including players who spent well over an hour actively playing, and who returned for more than one session on a second day. Merging the first two clusters beyond the low engagement level, and the top three clusters into larger clusters, we formed groupings of low, high, and very high engagement that make sense across play time, number of events per user, and number of sessions per user.

Table 3. Both D1 and D2 reflect that engagement is low overall relative to clusters on EI in this data. It is notable that D1 is session-based, thus, its output of 23% engagement means that 23% of sessions were engaged regardless of who was playing.

Engagement Definition	Engaged (%)	Very High Engagement (%)	High Engagement (%)	Not Engaged (%)	User- or Session-Based
D1	23.18	1.2	21.98	76.85	session
D2	19.8	n/a	n/a	80.2	user
EI	37.0	9.9	27.1	62.9	user

Grouped engagement levels using EI suggest that about 37% of users were engaged overall. D1 suggests that only about 23% of sessions were engaged, but with the EI groupings, we can say that of those engaged sessions, players with high EI were likely the ones performing them. Indeed, players with high EI values played greater numbers of sessions that lasted longer and tended to be more active compared with their peers. Furthermore, D2 may underrepresent engagement on a user level by a significant amount. A good engagement metric will accurately reflect engagement levels with respect to as many indicators of engagement as are measured in the data. Knowing the strengths and weaknesses of each definition, EI groupings shed light on engaged users that were lost in D2 and hidden beneath session-based metrics in D1. Likewise, EI groupings are highly correlated with three important indicators of engagement.

Limitations of EI at Present

EI has some potential weaknesses as well. In this case, using the moment of churn to calculate a probability of churn is intentionally simplified. It may also be that number of events aggregated per user would work well as a proxy for engagement in this particular case study. As we can see in Table 4, EI correlates conspicuously highly with the attribute reflecting number of events. This is not surprising, since the metric is event-based in this instance. EI as it stands for this case study treats all event types that are not log in or log out events equally. In reality, certain event types may be associated with longer, more active sessions, and some events will be associated with a higher risk of churn. Thus, the probability of churn (and the inverse probability of churn) can and should be updated to reflect reality in real-time, when the precise moment of churn is unknowable. In real-time, EI should correspond more strongly with behavior than with number of events alone. Regardless, number of events as well as length of play will likely always correlate highly with a solid engagement metric for an open world game.

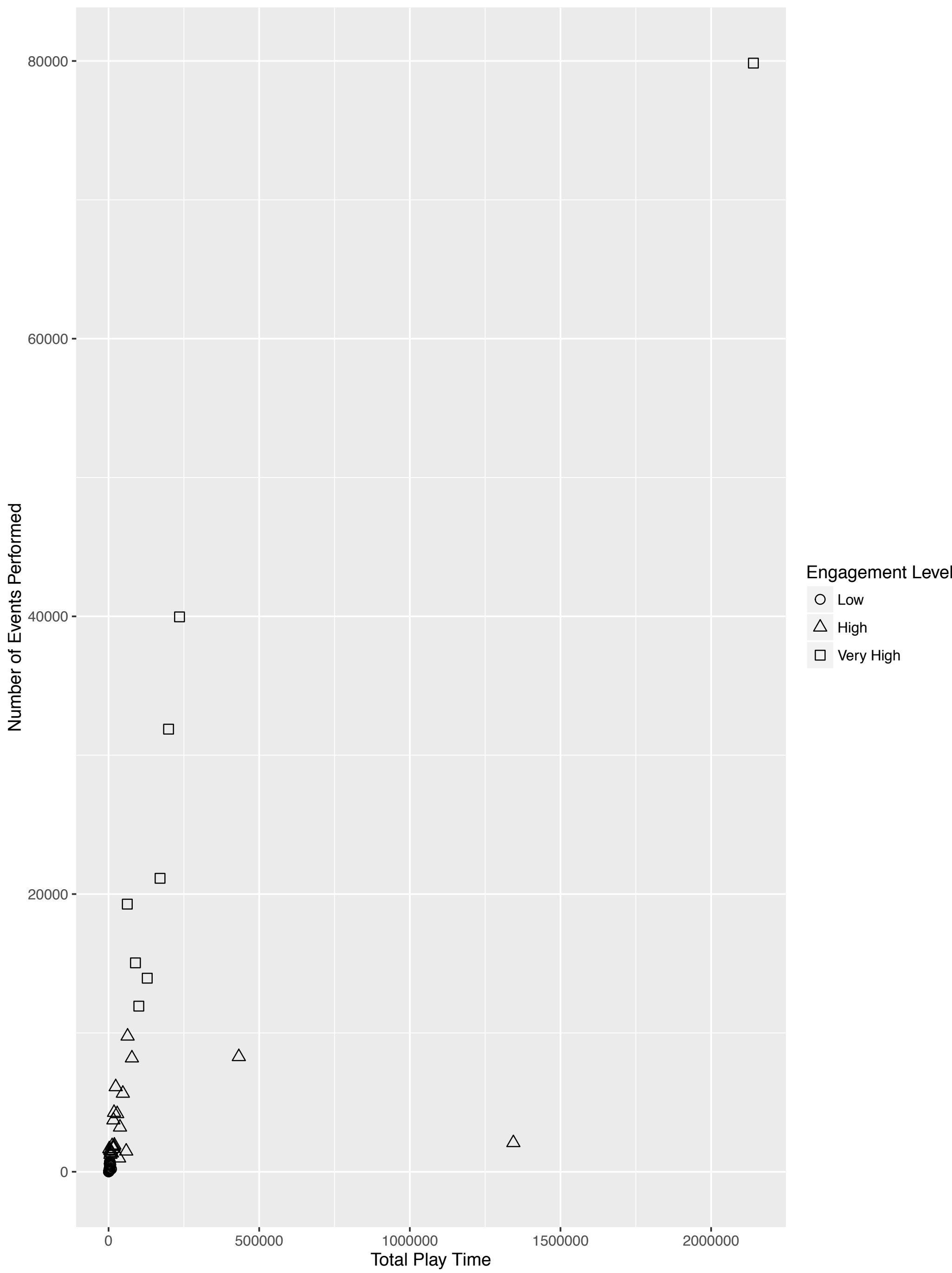

Figure 6 *K*-clusters on EI favor more active sessions over longer sessions. Even so, EI is strongly correlated with total play time and events per session (*see Table 4*).

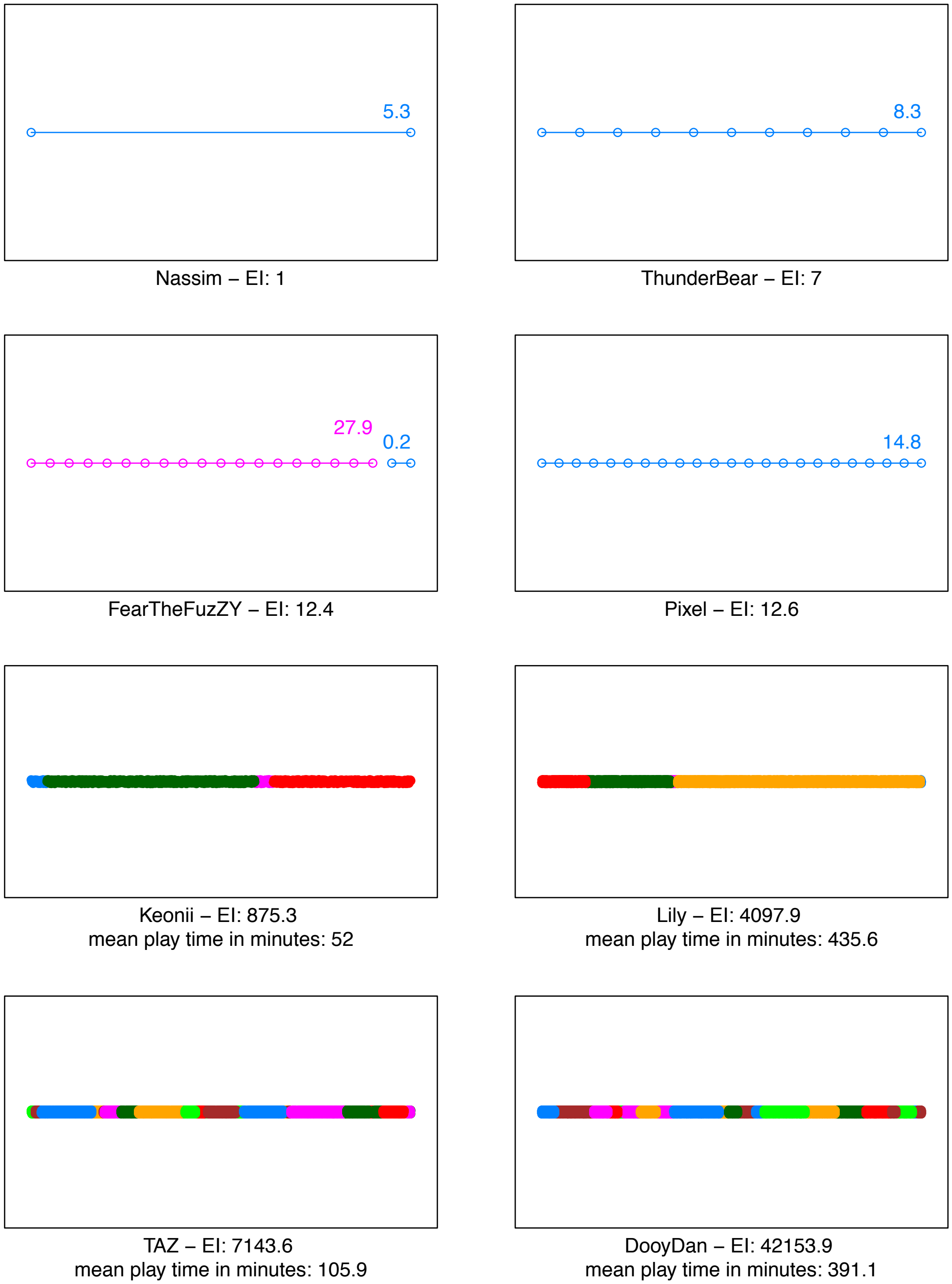

Figure 7 The above strip plots represent the events performed by a user. Distinct sessions are mapped to color. Labels on plots are actual playtime in minutes for the session labeled. Players with the lowest values for EI performed few events during their active sessions. Most players with low EI did not return for a second session. Players with high EI values played greater numbers of sessions that lasted longer and tended to be more active compared with their peers.

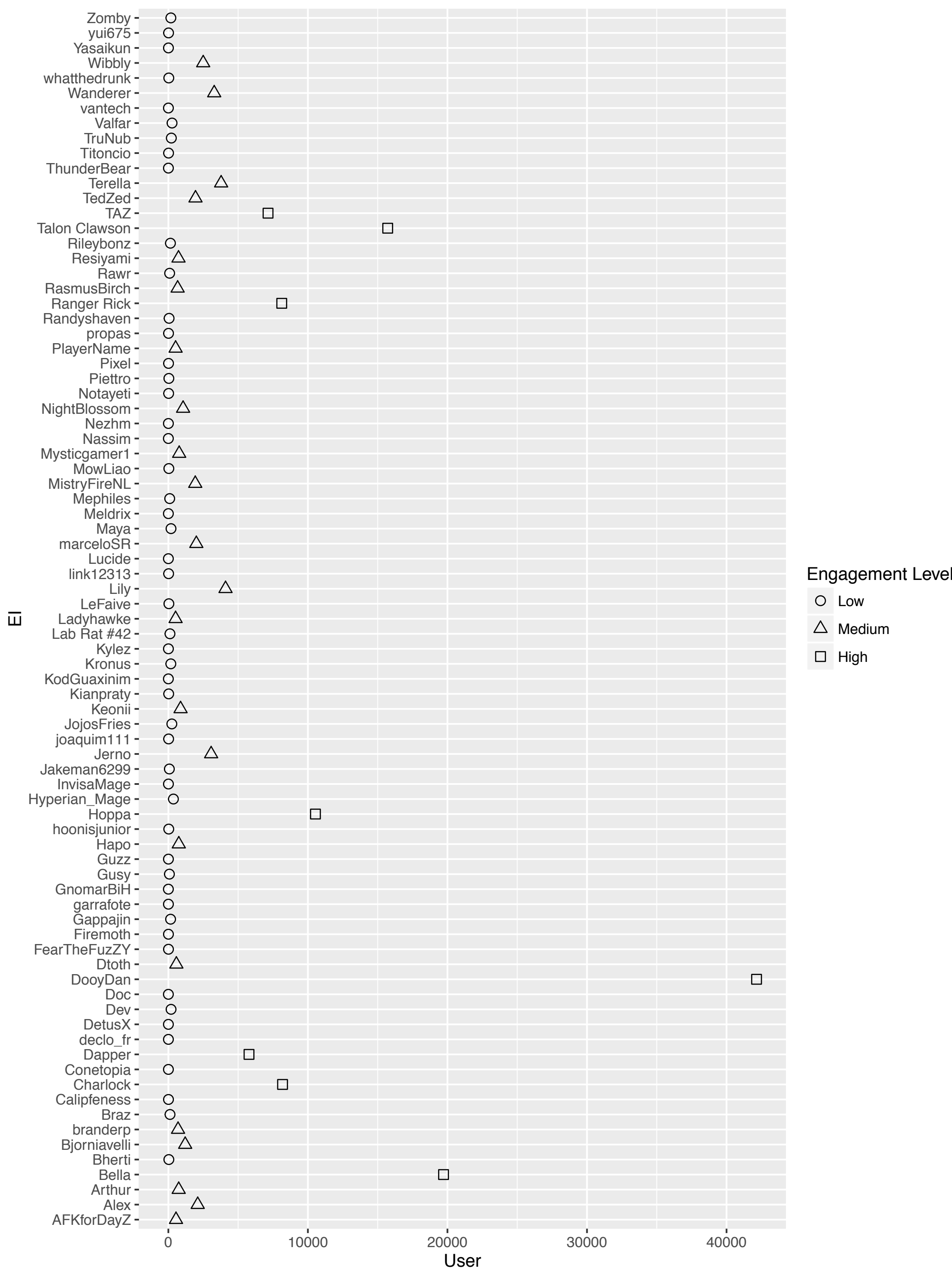

Figure 8 EI is highly correlated with events per session and total play time as well as number of sessions played by users (*see Table 4*). Low EI indicate low levels of activity by user. Above charts EI mapped to levels determined by *k*-means clustering.

Table 4 EI is more strongly correlated to all indicators of engagement than Definition 2. The weakest relationship for EI is with the total unique days a user played the game.

Correlations Across Attributes

	D2	**EI**	**Number of Events**	**Has Interacted**	**Total Playtime**	**Number of Sessions**	**Total Days**
Total Days	0.779	0.578	0.587	0.466	0.388	0.866	1.000
Number of Sessions	0.718	0.737	0.741	0.481	0.409	1.000	0.866
Total Playtime	0.441	0.755	0.742	0.141	1.000	0.409	0.388
Has Interacted	0.493	0.254	0.277	1.000	0.141	0.481	0.466
Number of Events	0.585	0.998	1.000	0.277	0.742	0.741	0.587
EI	0.564	1.000	0.998	0.254	0.755	0.737	0.578
D2	1.000	0.564	0.585	0.493	0.441	0.718	0.779

Defining Interaction

We were not able to augment the data to define social interaction in *TUG*. Some of the players from the beta test have been active users in the Nerd Kingdom forum where the progress of the game development is continually discussed (though the forums have become less active since the beta test). Nerd Kingdom was not able to survey every player from the beta test, nor to verify if users from the forum were the players from the game since user names for each were not linked in any way. Because we only had one option, defining user interaction was straight forward. We could not speculate if users actually met up on purpose, communicated, or even noticed each other during game play. The only way to determine if players interacted with one another was to see if they were near the same place at the same time. In other words, in this case, to say players interacted is more accurately to say that they had the opportunity to do so. To be more specific, proximity was used a substitute for true interaction in our definitions. While proximity and interaction have different meanings—interaction implies a host of qualities that, like engagement, are difficult to pin down quantitatively—the opportunity to interact is the extent to what we can demonstrate given the data that we have. For this study, we define a *gathering* as the phenomenon of any group of *n* players within 40 game units of each other within the time span of a minute. A user who has ever participated in a *gathering* is known as a *gatherer*. In the context of the above description of a *gathering*, information about *gatherers* can be summarized as follows.

There were 24 *gatherers* out of the total 82 active users. *Gatherers* were online for 40 out of the 42 days in range for the beta test data. The first *gatherer* signed on to *TUG* on January 8, 2016, two days after the beta test began. Actual *gatherings* occurred over a span of 17 days, where the first *gathering* occurred on January 28, 2016, 22 days after the launch of the beta test, and the last occurred February 14, 2016, three days before the final day of the beta test. Out of those 17 days, *gatherings* actually took place on only nine unique days. On five of those nine days, only one *gathering* took place. Of the remaining four days, there were four *gatherings* on January 28, two on February 1 and February 10, and a maximum of seven *gatherings* on February 2, 2016. Of the 24 unique *gatherers*, five participated in more than one *gathering*.

Among the five *gatherers* who *gathered* more than once, only one, username TAZ, *gathered* with two different *gatherers*. The other four could be grouped into *gatherer* pairs, as they only *gathered* with each other multiple times. Username TruNub and Username Jerno *gathered* five times on February 2, and username Resiyami and username Hapo *gathered* four times on January 28.

Results

How did *gathering* affect engagement level for players of *TUG*? As mentioned above, only 24 unique players ever *gathered*. Of these, 19 of these *gatherers* only did so once. Of the five *gatherers* whoever did so two or more times, only one was not classified as engaged. Indeed, all but three *gatherers* were not considered engaged. These results are promising, but there is not enough data to make statistically significant conclusions. We have enough information to make intelligent observations, however. For example, the

Bayesian breakdown of conditional probabilities from these two attributes reveals patterns that may imply a trend.

Table 5 The probabilities associated with the given states suggest that gathering may increase the likelihood of engagement or that the state of engagement may increase a player's chances of having gathered at some point.

Given:	**The Probability of:**	**P(A|B):**
gathered	engaged	87.5%
	not engaged	12.5%
did not *gather*	not engaged	84.0%
	engaged	16.0%
engaged	*gathered*	70.0%
	did not gather	30.0%
not engaged	did not *gather*	94.0%
	gathered	6.0%

Given a player *gathered*, she had a 87.5% probability of being engaged. Likewise, given a player did not *gather,* her risk of being engaged was a low 16%. Furthermore, there is was a 70% chance a player *gathered* if it was given that she was classified as engaged, and there was a low risk of 6% that a player *gathered* if it was given that she was not classified as being engaged. These results support the claim that the opportunity to interact in the game could have a positive effect on engagement. Thus we see that gathering may increase the likelihood of being engaged. The state of being engaged may also increase a players chances of having *gathered.* Certainly in the beta test, 21 of the 30 players classified as engaged were *gatherers*.

A logistic regression model provides us with another way to communicate seeming trends in conditional probabilities, since the log odds the model is built on are closely related to the conditional probabilities for our predictor and target events. For example, a logistic regression model predicts that engaged users (those who were in the top two EI clusters) *gathered* with an accuracy of 85.7% (Equation 3). The state of being engaged is a statistically significant predictor below the 0.001 level. The model fit is substantial. The gap between residual deviance and null deviance is 32.9 units in favor of the residual, and McFadden's R^2 is 0.42. A similar logistic regression model predicts that *gatherers* were engaged with 85.7% accuracy as well (Equation 4).

$$\text{(3) } y^* = -2.741 + 3.791(x)$$

$$\text{(4) } y^* = -1.488 + 3.791(x)$$

These models reinforce what we see in Table 5; in particular, Equation 4 demonstrates that if *x,* whether or not a player has *gathered*, is true, then the likelihood that the player fell into a high engagement category was 91%. Equation 3 similarly demonstrates that if a player fell into a high engagement category, that player was 74% likely to have *gathered.*[1] From this we can see that the act of *gathering* and the state of engagement are overlapping states. It may be reasonable to suggest that one entails the other.

That said, it is important to select good definitions of engagement and interaction based on solid methodology. Selecting a stricter definition of engagement with a smaller percentage of engaged users, like D2, for example, can flip conditional probabilities in ways that suggest that the odds of both gathering and of being engaged are too low to expect either to happen very often. Even so, in the current model, we

[1] Recall that logistic regression models return linear equations which represent the log odds of an event *y.* To determine the probability of *y* given the log odds of *y,* we use $p(y) = e^{y^*}/(e^{y^*} + 1)$.

must acknowledge that our definition of interaction is intentionally open-ended due to the information that was available. Results indicate that players who were less likely the churn had more opportunities to interact than other players who did not spend a lot of time in the game. We cannot conclude that proximity to other players strengthened an engaged state, or reduced a player's likelihood of churn. However, we can claim that an engaged state did considerably increase the likelihood that players would find themselves in close proximity to other players. Further, we have sufficient evidence to propose that *gathering* may be a strong indicator of engagement itself.

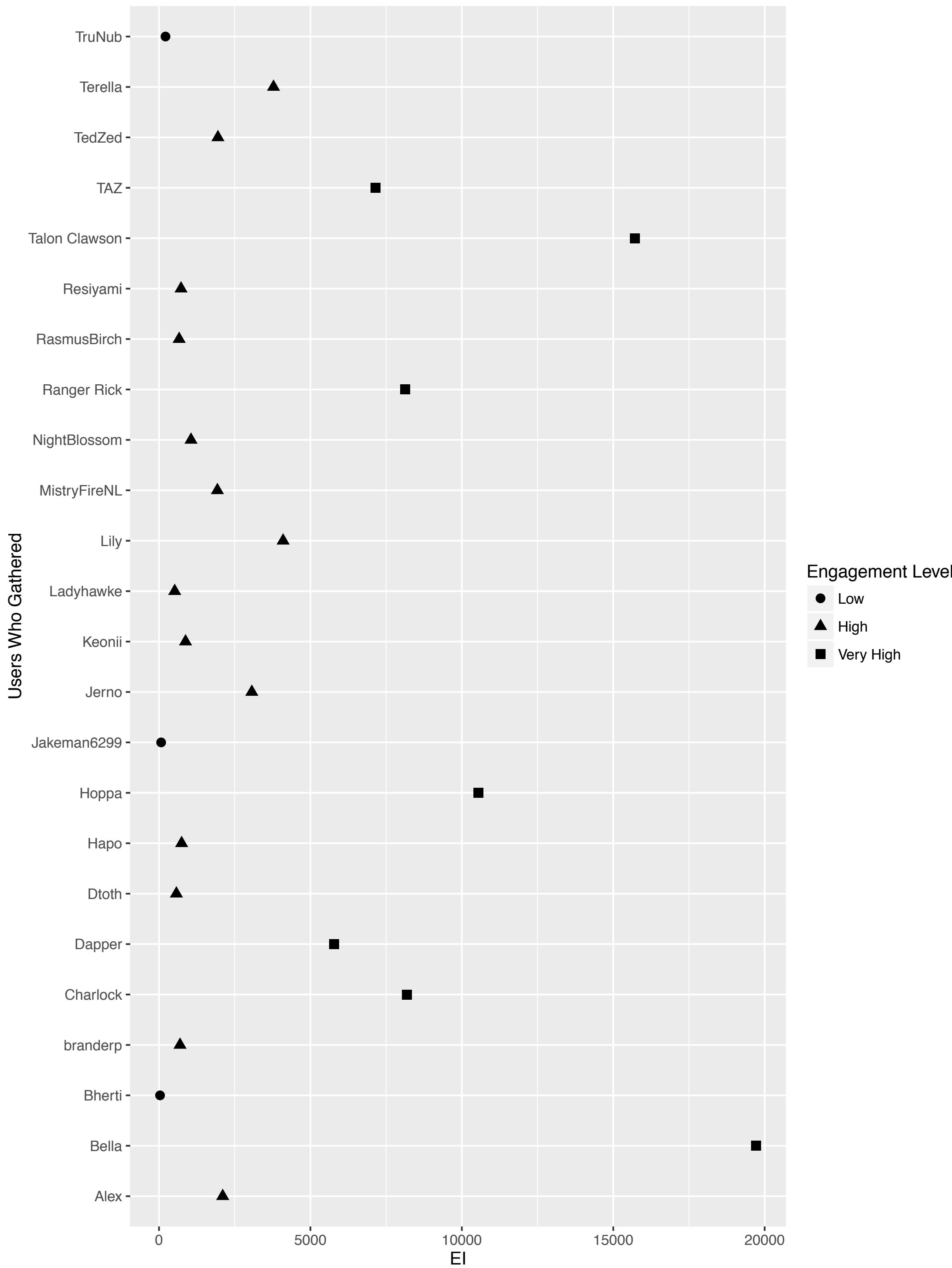

Figure 9 The same chart as shown in *Figure 8* featuring players who *gathered* only. Only three *gatherers*, 13% of all *gatherers*, had low enough EI values to be classified with low engagement.

Chapter 5: Conclusion and Future Work

While it may not be prudent to conclude that *gathering* affects engagement in *TUG*, we know that the most active players who spent the most time in the game also tended to *gather.* Indeed, *gathering* may be an indicator of engagement in *TUG.* By our definition of engagement, this statement implies that players who did not *gather* were more likely to churn. In the future, analysis of events that increase a player's likelihood of churn should be studied in greater depth. Determining more realistic probabilities of churn associated with event types and other indicators we can harvest from player logs, such as *gathering* itself, will allow us to strengthen EI. Further, applying EI to larger data sets will allow us to test and verify its validity as an engagement metric. In particular, data from the official release of *TUG* would be the most suitable. Likewise, it would be useful for methodological discussion regarding analytical approaches to measure engagement to expand to a range of other sandbox games. This study was focused on avoiding qualitative methods such as surveys and interviews for pragmatic reasons, but these methods, would allow us to build a much stronger definition for the state of interaction. Furthermore, qualitative methods would benefit future analyses if used to validate and refine analytic definitions for the state of being engaged as well. It would be useful to test and refine the concept that engagement can be represented accurately in quantitative terms as a kind of inversion of churn.

References

Attfield, S., Kazai, G., & Lalmas, M. (2011). Towards a science of user engagement (Position Paper). *WSDM Workshop on User Modelling for Web Applications*. Retrieved from http://www.dcs.gla.ac.uk/~mounia/Papers/engagement.pdf

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., Mcbroom, E., Burkhart, K. M., & Pidruzny, J. N. (2009). Journal of Experimental Social Psychology The development of the Game Engagement Questionnaire : A measure of engagement in video game-playing. *Journal of Experimental Social Psychology*, *45*(4), 624–634. http://doi.org/10.1016/j.jesp.2009.02.016

Castronova, E. (2006). On the Research Value of Large Games: Natural Experiments in Norrath and Camelot. *Games and Culture*, *1*(2), 163–186. http://doi.org/10.1177/1555412006286686

Denisova, A., Nordin, A. I., & Cairns, P. (2016). The Convergence of Player Experience Questionnaires. *CHI PLAY '16: The 2016 Annual Symposium on Computer-Human Interaction in Play*, 33–37. http://doi.org/10.1145/2967934.2968095

Drachen, A., & Canossa, A. (2011). Evaluating motion: spatial user behaviour in virtual environments. *International Journal of Arts and Technology*, *4*(X), 294. http://doi.org/10.1504/IJART.2011.041483

Drachen, A., & Schubert, M. (2013). Spatial game analytics and visualization. In *IEEE Conference on Computatonal Intelligence and Games, CIG*. http://doi.org/10.1109/CIG.2013.6633629

Drachen, A., Sifa, R., Bauckhage, C., & Thurau, C. (2012). Guns, swords and data: Clustering of player behavior in computer games in the wild. In *2012 IEEE Conference on Computational Intelligence and Games, CIG 2012* (pp. 163–170). http://doi.org/10.1109/CIG.2012.6374152

Drachen, A., Thurau, C., & Bauckhage, C. (2013). A Comparison of Methods for Player Clustering via Behavioral Telemetry (Pre-print). *Foundations of Digital Games 2013*.

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J. (2006). "Alone together?": exploring the social dynamics of massively multiplayer online games. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '06*, (March), 407. http://doi.org/10.1145/1124772.1124834

Duggan, M. (2015). Gaming and Gamers. Retrieved from http://www.pewinternet.org/2015/12/15/gaming-and-gamers/

Eastin, M. S., Daugherty, T., & Burns, N. M. (2010). *Handbook of research on digital media and advertising: User generated content consumption*. *Handbook of Research on Digital Media and Advertising: User Generated Content Consumption*. http://doi.org/10.4018/978-1-60566-792-8

ESA. (2017, January 17). U.S. Video Game Industry Generates $30.4 Billion in Revenue for 2016. Retrieved April 1, 2017, from http://www.theesa.com/article/u-s-video-game-industry-generates-30-4-billion-revenue-2016/

Kawale, J., Pal, A., & Srivastava, J. (2009). Churn prediction in MMORPGs: A social influence based approach. In *Proceedings - 12th IEEE International Conference on Computational Science and Engineering, CSE 2009* (Vol. 4, pp. 423–428). http://doi.org/10.1109/CSE.2009.80

Kirschner, D., & Williams, J. P. (2015). Measuring Video Game Engagement Through Gameplay Reviews. *Simulation & Gaming*, *45*(4–5), 593–610. http://doi.org/10.1177/1046878114554185

Kwok, M., & Yeung, G. (2005). Characterization of user behavior in a multi-player online game. *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, 69–74. http://doi.org/http://doi.acm.org/10.1145/1178477.1178486

Lehmann, J., Lalmas, M., Yom-Tov, E., & Dupret, G. (2012). Models of User Engagement. In J. Masthoff, B. Mobasher, M. C. Desmarais, & R. Nkambou (Eds.), *User Modeling, Adaptation, and Personalization: 20th International Conference, UMAP 2012, Montreal, Canada, July 16-20, 2012. Proceedings* (pp. 164–175). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-31454-4_14

Monson, M. J. (2012). Race-Based Fantasy Realm: Essentialism in the World of Warcraft. *Games and Culture*, *7*(1), 48–71. http://doi.org/10.1177/1555412012440308

Muller, S., Frey, S., Kapadia, M., Klingler, S., Mann, R. P., Solenthaler, B., … Gross, M. (2015). HEAPCRAFT: Quantifying and Predicting Collaboration in Minecraft. In *Artificial Intelligence and Interactive Digital Entertainment* (pp. 156–162).

Nardi, B., & Harris, J. (2006). Strangers and Friends: Collaborative Play in World of Warcraft. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work* (pp. 149–158). New York, NY, USA: ACM. http://doi.org/10.1145/1180875.1180898

Needleman, S. E. (2017). How Mobile Games Aim to Keep You Coming Back. Retrieved from https://www.wsj.com/articles/how-mobile-games-aim-to-keep-you-coming-back-1483790404

Nerd Kingdom. (2013). TUG by Nerd Kingdom - Kickstarter. Retrieved April 1, 2017, from https://www.kickstarter.com/projects/nerdkingdom/tug-1

Newzoo. (2016). THE GLOBAL GAMES MARKET REACHES $99.6 BILLION IN 2016, MOBILE GENERATING 37%. Retrieved April 1, 2017, from https://newzoo.com/insights/articles/global-games-market-reaches-99-6-billion-2016-mobile-generating-37/

O'Brien, H. L., & Toms, E. G. (2008). What is user enagement? A Conceptual Framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, *59*(6), 938–955. http://doi.org/10.1002/asi.20801.1

O'Brien, H. L., & Toms, E. G. (2010). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*. http://doi.org/10.1002/asi.21229

Pirker, J., Rattinger, A., Drachen, A., & Sifa, R. (2016). *Analyzing Player Networks in Destiny*. Retrieved from https://andersdrachen.files.wordpress.com/2016/11/playernetworks.pdf

Raimbault, M. S., Clark, C., & Nicholson, C. D. (2016). *Behavioral Clustering in Open World Sandbox Games*. Southern Methodist University.

Schoenau-Fog, H., & Henrik, S.-F. (2014). The Player Engagement Process – An Exploration of Continuation Desire in Digital Games. *DiGRA 2011 Conference: Think Design Play.*, *6*, 1–18. Retrieved from http://www.digra.org/digital-library/publications/the-player-engagement-process-an-exploration-of-continuation-desire-in-digital-games/%5Cnhttp://www.digra.org/wp-content/uploads/digital-library/11307.06025.pdf

Wiebe, E. N., Lamb, A., Hardy, M., & Sharek, D. (2014). Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. *Computers in Human Behavior*, *32*, 123–132. http://doi.org/10.1016/j.chb.2013.12.001

Yee, N. (2006). Motivations for Play in Online Games. *CyberPsychology & Behavior*, *9*(6), 772–775. http://doi.org/10.1089/cpb.2006.9.772

Yee, N., Ducheneaut, N., Nelson, L., & Likarish, P. (2011). Introverted Elves & Conscientious Gnomes: The Expression of Personality in World of Warcraft. *Proceedings of the Annual Conference on Human Factors in Computing Systems - CHI '11*, 753–762. http://doi.org/10.1145/1978942.1979052

Zhuang, X., Bharambe, A., Pang, J., & Seshan, S. (2007). *Player Dynamics in Massively Multiplayer Online Games*. Carnegie Mellon University, Pittsburg, PA.