

SAS[®] GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL

**Creating Personal Game Statistics for Video Game
Events**

Examples from EVE Online and Ingress

USERS PROGRAM



Creating Personal Game Statistics for Video Game Events: Examples from EVE Online and Ingress

Peter Timusk

Statistics Canada

[ABSTRACT ICLICK TO EDIT\)](#)

Computer and video games are complex these days. Events in video games are in some cases recorded automatically in text files, creating a history or story of game play. There are countable items in these event records that can be used as data for statistics and other types of modeling. This E-Poster shows you how to statistically analyze text files for video game events using SAS®. Two games are analyzed. EVE Online, a massive multi-user online role-playing spaceship game, is one. The other game is Ingress, a cell phone game that combines exercise with GPS and real-world environments. In both examples, the techniques involve parsing large amounts of text data to examine recurring patterns in text that describe events in the game play.

[METHODS](#)

- **Data Input:**
- In the case of EVE Online game, we took approximately 500,000 text files, each about 1 kilobyte in size and combined these into a SAS dataset with only one text variable covering the width of the file.
- In the case of the Ingress game with Gmail text files, the archiving process at Gmail allowed an 'mbox' text file to be made that had all the game e-mail notifications already gathered into one file. There were approximately 2300 notification emails in this data. Again this was converted to a SAS dataset with one character variable covering the width of the text file data.

[METHODS](#)

Text mining was very simple for the EVE Online game. We made a date time variable from each small log file where the first 17 characters are a date time stamp, as is the name of the file. This is completed with a complex regular expression using the PRXMATCH function and the SUBSTR function. The result can be formatted to a SAS date time variable using an input function.

Here is the code:

```
gatedatetime=input((trim(substr(filename,prxmatch('/([[:digit:]]{4})([[:punct:]]{1})([[:digit:]]{2})([[:punct:]]{1})([[:digit:]]{2})([[:blank:]]{1})([[:digit:]]{2})([[:punct:]]{1})([[:digit:]]{2})([[:punct:]]{1})([[:digit:]]{2})/','filename),(prxmatch('/([[:digit:]]{4})([[:punct:]]{1})([[:digit:]]{2})([[:punct:]]{1})([[:digit:]]{2})([[:blank:]]{1})([[:digit:]]{2})([[:punct:]]{1})([[:digit:]]{2})([[:punct:]]{1})([[:digit:]]{2})/','filename')+19))),YMDDTTM19.);
```

In addition, an action of the game play called 'jumping' was used as a filter. Every time the player of the game did 'jumping' it was recorded in the log file with the word 'jumping'. As this 'jumping' progresses, a player sees the notice 'jumping' on their screen. The notice is written to a log file in the player's 'My Documents' folder. Because the jumping note in the log file also includes geographic place names, called galaxies, one can draw a route map of the game play as the player jumps from galaxy to galaxy. In an industrial process like courier delivery or supply logistics, a time location stamp could be parsed in a similar way.

In a further analysis, we search the galaxies for game properties that indicate the player was taking unknown risks early in game play. This could be related to addictions theory, which is of concern in problem gambling and risk-taking behavior.

Creating Personal Game Statistics for Video Game Events: Examples from EVE Online and Ingress

Peter Timusk

Statistics Canada

METHODS

Text mining was far more complex for the Ingress game. We have one character variable containing all the emails and email fields in text format. To begin we search for the location (or observation in our dataset) of the beginning of every email by identifying occurrences of the email header which always contain the same text string. Marking the beginning of each email by the observation number in the SAS dataset allows a splitting of each email into a temporary SAS data set. This is then parsed for the date time stamp, the enemy agent's name and the location of the game piece in latitude and longitude values. The values, although they are true latitude and longitude values, will not show a game piece unless viewed on the game company's web site map. This 'Intel' map is the only way outside of the cell phone app itself to view the game pieces. As a consequence each latitude and longitude location needs to be encoded in a web address for the Intel map web site.

The common name of the game piece is also parsed out of the email. Each enemy agent attack is given its own email. If a different agent attacks, a different email notification is sent. If the same enemy agent attacks a different game piece again, a new different email notification is sent. Thus, it is reasonable to parse out only one agent's name and only one game piece from each email. This is not the full picture as attacking one game piece may affect other game pieces and the email will also mention these locations. For the present analysis, one enemy agent's name and one game piece location are parsed out.

A date time stamp is also parsed out. Technically, we have full location and time references here, although the z-axis (height) is not really needed.

RESULTS

We are able to examine game play for both games. For the EVE Online game, we can see when the new player took a risk by going into an area in the game board of higher risk. We can follow the route the player took.

For Ingress we have a statistics table of the sum of enemy attacks sorted by enemy agent names. To protect player's identities, the names have been changed to a random food name concatenated with a random animal name. However, the statistics are actually from the author's game play.

Table 1. Top ten enemy agent attack count and percent of total attacks.

EnemyAgent	COUNT	PERCENT
MouseApples	1,254	7.84
CatPeaches	1,187	7.42
PoultLemons	1,164	7.28
LionLemons	646	4.04
MousePineapples	542	3.39
LizardPeaches	478	2.99
ChickCherries	466	2.91
ChickBlueberries	436	2.73
ChickenPeaches	425	2.66
ChickPeaches	366	2.29

Creating Personal Game Statistics for Video Game Events: Examples from EVE Online and Ingress

Peter Timusk

Statistics Canada

CONCLUSIONS

We have shown that text mining can be completed in Base SAS with knowledge of the data and basic string-handling techniques. This project has shown that game data presents a source of data for analysis and further studies. Likewise, much event data from industry could be given similar treatment to allow analysis within industrial processes. If a date, time and location can be parsed out of event stream data, statistical analysis is possible that could provide meaningful results.



SAS[®] GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL