

# Gaming the System

*Anna Huynh*

GameFAQs is a website that hosts FAQs(frequently asked questions) and walkthroughs for video games. It also operates as a forum with message boards devoted to every video game and even non-video game related topics. One very important and possibly overlooked property that GameFAQs has is how it also serves as a video game database. A user is able to look up any video game title and get an idea on what the video game is like based on other users' experiences. GameFAQs' database is a very useful tool in helping a user decide whether they should play the game or not.

Because there is not enough time in a human's lifespan to play every possible video game, we use K-Means clustering to help allow us to decide which video games are worthwhile. The first step in even doing so is scraping data from GameFAQs. However, the scraping process is not that simple with GameFAQs. They claim to prefer that  $<5$  requests are sent each second, but will still block a user for sending a request every 5-10 seconds and even every 30-45 seconds. The optimal time frame in sending a request ended up being every 60-90 seconds. Unfortunately, this made acquiring the data a very time-consuming process.

Because of time constraints, only RPG's(role-playing games) for the PSP, DS, SNES, PS1 and PS2 were considered. In every RPG's game page, we extracted the following information:

Game title - obvious.

Number of owners - number of GameFAQs users who claim to own the game.

Rating - average rating, which is determined by GameFAQs users. The scale is out of 5.

Numbers of owners rated - number of GameFAQs users who claim to own the game and also gave said game a rating.

Difficulty - the difficulty of the game which is determined by a popular vote. The difficulty ratings are "Easy", "Easy-Just Right", "Just Right", "Just Right-Tough", "Tough", "Tough-Unforgiving" and "Unforgiving".

Percentage of difficulty ratings - percentage of votes that rated the game the listed difficulty.

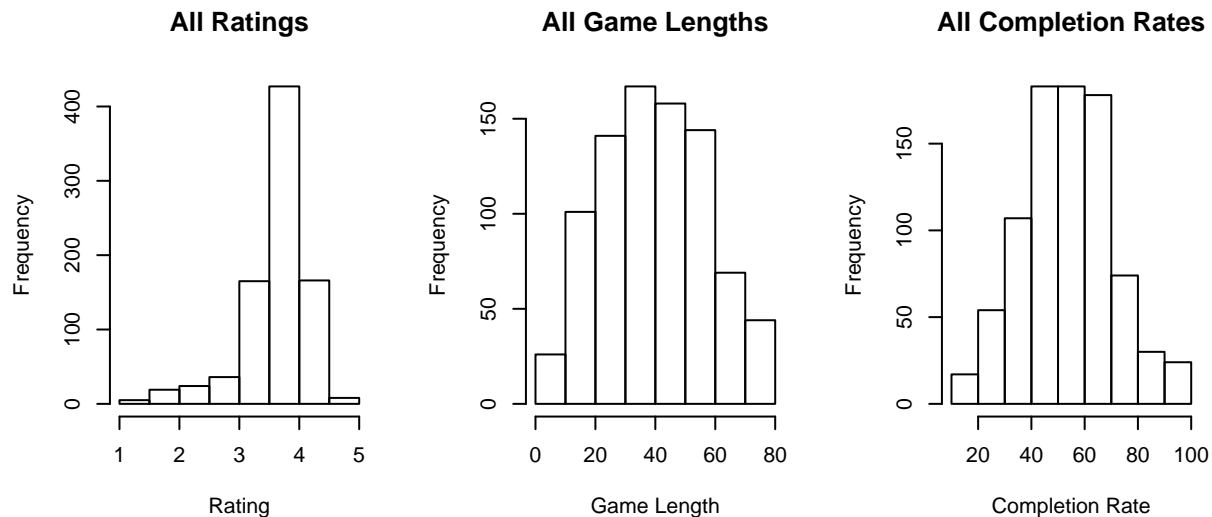
Game length - average game length, which is determined by GameFAQs users. The unit is in hours.

Votes for game length - number of GameFAQs users who gave how long it took for them to beat the game.

Number of users who completed the game - obvious.

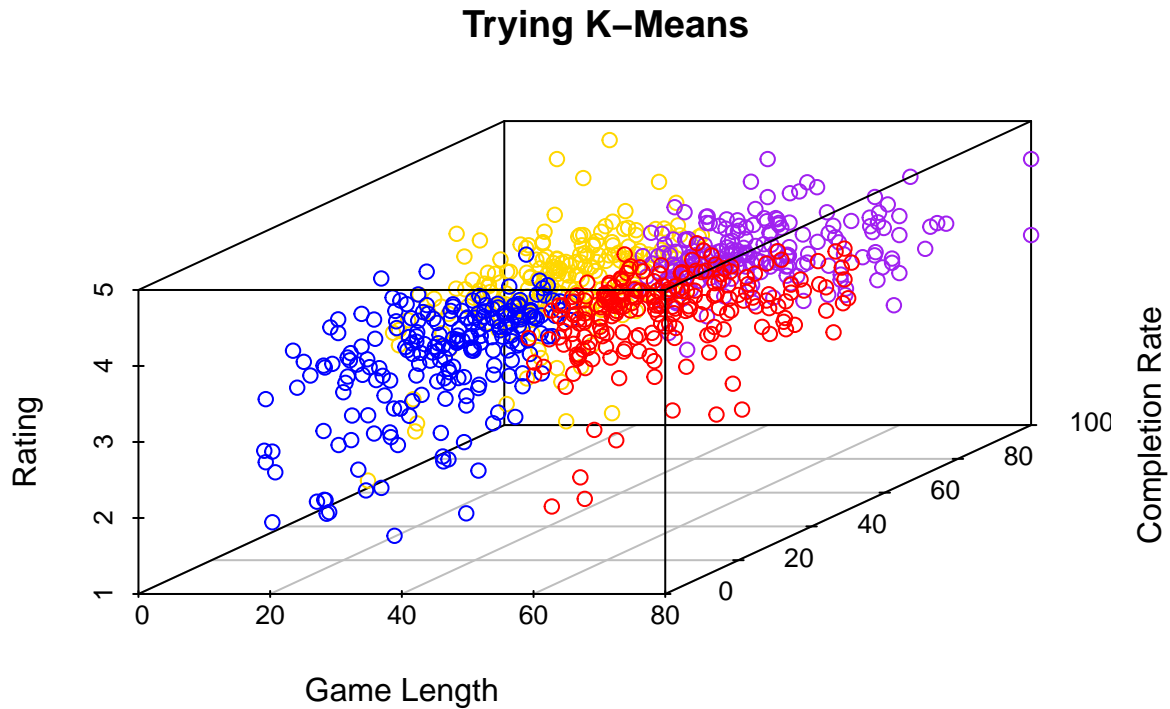
Completion rate - determined by dividing the number of users who completed the game by the number of owners.

After acquiring all the relevant game information, our next step was making sure that our variables make sense. Because we are interested in using K-Means clustering, we mainly focus on looking at the numerical variables.



Luckily, the numerical variables do make sense. An RPG's game length can vary heavily depending on how much content it has. The bulk of the ratings lingering between 3 and 4 also makes sense since games that are extremely spectacular or extremely terrible are harder to come by. Completion rates also having a wide spectrum makes sense because it does take a bit of commitment to beat an RPG due to their relatively long game lengths compared to other game genres.

We then apply K-Means clustering with these three numerical variables and acquire the following plot:



The most appealing cluster appears to be the top right one: the games have fairly high ratings, fairly high completion rates, and game lengths that are enough to hook the gamer for a while. Some of the games in the cluster are:

```
## [1] "Summon Night EX-Thesis: Yoake no Tsubasa"
## [2] "Kingdom Hearts"
## [3] "Eiyuu Densetsu: Ao no Kiseki"
## [4] "Shinseiki Odysseya 2"
## [5] "Tales of Symphonia"
## [6] "Atelier Iris 2: The Azoth of Destiny"
## [7] "Lagnacure"
## [8] "Dragon Quest V: Hand of the Heavenly Bride"
## [9] "TwinBee RPG"
## [10] "Radiata Stories"
## [11] "Chrono Cross"
## [12] "Spectral Force: Radical Elements"
## [13] "Kingdom Hearts II"
## [14] ".hack//Fragment"
## [15] "Pokemon Mystery Dungeon: Explorers of Time"
## [16] "Khamrai"
## [17] "Aretha II: Ariel no Fushigi na Tabi"
```

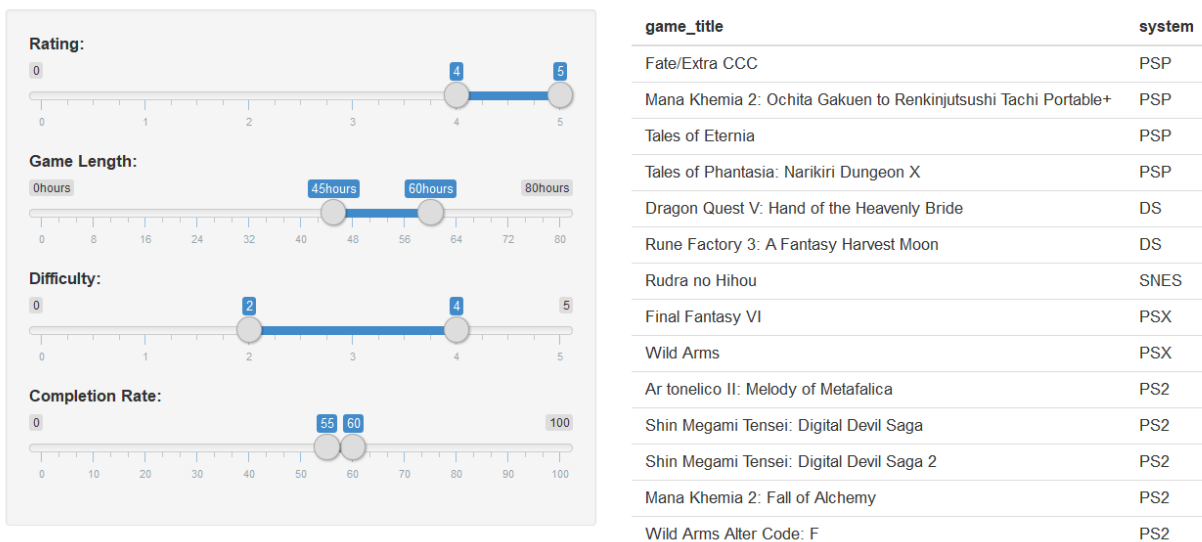
```
## [18] "Final Fantasy V"
## [19] "Nishikaze no Kyoushikyoku: The Rhapsody of Zephyr"
## [20] "UnchainBlades EXXiV"
```

Based on our experience with RPG's, we would say that the K-Means cluster did a pretty good job grouping the RPG's together.

Lastly, we made a new numerical difficulty variable by converting the difficulty ratings to numerical values on a scale of 5. 2/5 would be "Easy", 2.5/5 would be "Easy-Just Right" and so on. This variable along with the rating, game length, and completion variables would be used for our Shiny App, which can be accessed at [https://ripsilon.shinyapps.io/shiny\\_app/](https://ripsilon.shinyapps.io/shiny_app/).

The goal of our Shiny App is to tell a user what RPG's they should try out depending on what they are looking for in an RPG. Below is an example of us using the application.

## Criteria for the RPG's you want to play



Again, based on our experience with RPG's, we would say that the Shiny App does a good job of filtering out RPG's based on a user's interests.