**R project proposal**

Haonan Guo, Janet Zou, Jintao He, Xiazijing Zhong

Description of the situation:

Assuming a publisher is developing a new game to be released in North America(NA) market, we want to know what game genre is most preferred, and on what platform would attract more sales in the NA market. In addition, we plan to investigate competitors’ regional focus. We plan to analyze 11562 different games along with their genre, publisher, regional sales and ratings. In the end, we aim to use a heatmap to present us the result of the most preferred game genre and platform in the NA market.

Data Source:

The data source we obtained is from the Kaggle website, and is available for public use. The dataset is in the form of one csv file, it contains 11562 lanes of different games and 16 columns of different variables(name, platform, release year, genre, publisher, sales based on markets: NA, EU, JP, Other and Global, critic and user’s count and score). We’ll use the readr library to read the dataset into a data frame in R, then clean out the unwanted variables and analyze the data from there. <https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings>

Question to be Answered with Data:

Which video game platform is most successful, in terms of the game bases and NA sales? And which platform is recommended in the trend? How is genre preference (user\_score) in North America? Do publishers have a regional focus (JP sales vs NA sales)? What are the top 10 video game IPs(intelligent property: based on ratings and sales) in the North America market? (Advanced Question)