

# DataFest 2022

TEAM 5 - Ken Vu, Kotomi Oda, Zhaoshan Duan

April 24, 2022

# Introduction

For this data set, we're working with the file `logs.csv`, which tracks all of actions enacted by each player during the game. Our work on this data set will focus on the following:

## Research Question

What are the **top three** types of events that took up the most playing time for players from each school?

## Goal

- ▶ Prepare the data to make it ready for analysis
- ▶ Address and explore our research question using our data

# Data Processing

- ▶ Got four players with some timestamps that failed to be generated
- ▶ One player had faulty timestamps for all their logs, thus they were removed from the analysis.
- ▶ NEW METRIC: `time_spent`, the amount of time player spent on the event.
  - ▶ first sorting the logs in ascending order by *player\_id* and *event\_time\_dbl*.
  - ▶ Then, for each player, we took the time difference (in seconds) between the current event and the next event.
  - ▶ That difference is saved to a new column called *time\_spent*
- ▶ Re-coded factorial levels of the schools from A to L.

# Result - Final Visualization

- Across all schools, the top three event types taking up the most playing time are **Challenge Stack**, **Knowledge Minigame**, and **People Sense**.

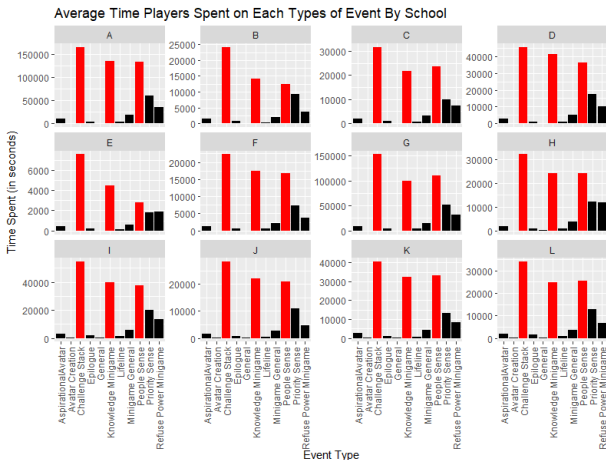


Figure 1: Note “average” here refers to median

## Conclusion - Future Research Opportunities

- ▶ ANOVA (Analysis of Variance) for each school to see if the median time spent is significantly different across all event categories
- ▶ Explore how much time each player is spending on each of the stacks in the Challenge Stack