PUBG Data Analysis

Team Members:
Phi Trang, Neil Nguyen, Jake Tran

Description

We plan to use the data to:

- Analyze most common player kill/death locations
- Find the average distance each weapon is used at
- Find correlation between distance traveled and rank
- Find the most popular weapons correlated to time in game
- Determine the weapon of choice and location to win a game

Prior Work

- A heat map was created to see where most final circles are
- A heat map was created to see where most kills occur
- A heat map of where most people jump out of the plane

Datasets

- Aggregate and deaths data scraped from pubg.op.gg
- Over 65 million death entries for PlayerUnknown Battlegrounds
- https://www.kaggle.com/skihikingkevin/pubg-match-deaths
- The dataset is downloaded on Jake’s workstation
- A sample of the dataset on GitHub

Proposed Work

- Data Cleaning
  - Sort the data by map, only include necessary attributes like weapon, killer, and victim information
  - Select a subset of the data because 65 million objects might be too much

List of Tools

- Anaconda/Python3
- Jupyter Notebook
- Pandas and Numpy
- Matlab Plot

Evaluation

We can evaluate our results by:

- Creating and plotting relationships between variables such as kill/death locations
- Plotting relationships between type of gun and distance of bullet travelled
- Finding correlation between distance travelled and end game rank
- We will run statistical analysis and check whether each correlation is significant or not