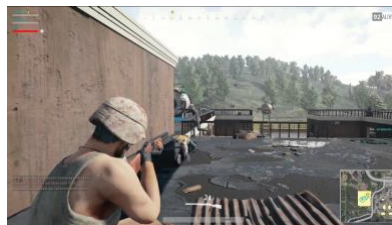# PUBG Visualisation

ZHERAN XU

PS4296875

Introduction

What's PUBG?

- PlayerUnknown's Battlegrounds (PUBG) is a popular multiplayer online battle royale game.

- Around 100 players parachute to the field, find weapons and other equipment from nothing to, and then fight to the last team while enter the gradually shrinking safe zone.

Data Description

Data Source

- The data set is crawled from "pubg.op.gg", which has match records. The data records a professional players solo ladder in 2018 Feb, where there is 31 matches in total.

- Each match has two JSON files and they have been transferred into csv as deaths and stats data set.

- In order to generally analyse the players' behaviors, the data sets are merged altogether.

Data

● Deaths

Variebles	Description	Datatype	e.g.
map_id	Playground map	Nominal	MIRAMAR
time_event	Seconds since match start	interval	801
description	The weapon or other non-combat factor	Nominal	M416
victim_name	Players ID who is knocked out	Nominal	huntingJJ
victim_x	Map X axis value	interval	642213.4
victim_y	Map Y axis value	interval	290567
victim_rank	Players' rank in this match	Ordinal	37
killer_name	Players ID who knocked out the opponent. (Could be empty)	Nominal	Yang_0830
killer_x	Map X axis value (Could be empty)	interval	651323.3
killer_y	Map Y axis value (Could be empty)	interval	305027.4
killer_rank	Players' rank in this match (Could be empty)	Ordinal	13

● The deaths information generally reflects the fight happened between players, although players may die for non-combat factor such as falling.

● Each case involves the basic relation:

◆ Player(victim) is knocked out by other player(killer) in different weapon(description), at somewhere(locations) and sometime(time_event)

◆ Player(victim) is knocked out by non-combat reason, at somewhere(locations) and sometime(time_event)

◆ there are two types of map in the match, and each of them is 800000x800000, while the location could be determined by '_x' and '_y' value.

- Stats

Variebles	Description	Datatype	e.g.
total_rank	Total players number in one match	Ordinal	88
nickname	Players ID	nominal	shazhuzhezhe
rank	Players final rank	ordinal	41
kills	The number of opponents knocked out by players	ratio	1
damage_dealt	The number of opponents knocked out by players	Ratio	98
walk_distance	The distance that player move on foot.	Ratio	526.558044
ride_distance	The distance that player move by vehicles.	Ratio	0
time_survived	Player survived times from the start of the match.	Ratio	1167.535

- As for the stats information, it mainly indicates the performance of each players in the match, such as players fight performance, march distance and final rank.

Question Description

Initial Questions

Q1: Map - Where does the battle occurs? And which areas are more intensive?

- Data Manipulation

  I introduce the game by explaining prototype to potential users, who are most of my friends and classmate, and ask their general idea about game design. After having a basic concept of game, I discuss more about different aspects - trigger setting / area control / exhibits related / seam through the journey.

- Visual Encoding

  Plot deaths cases as Scatter on the map image by position (x,y) , using hue to illustrate the time attribution.

ERANGEL   MIRAMAR

- Analysis and Answers
  - At the early period of game, most of the fights happen at the white and red circle area, where usually have a name on the map. Places in red circle, are considered to be heated battle field as more intensive scatters
  - As match goes on, the scatters widely spread but mainly in the orange circle, which is covers most centre area of map. It indicates that players move out of the initial location.

Q2: How does player die in the match? What kind of the weapons complete most part of kills?

According to game rules, the description of death could be divided into factor groups as follow:
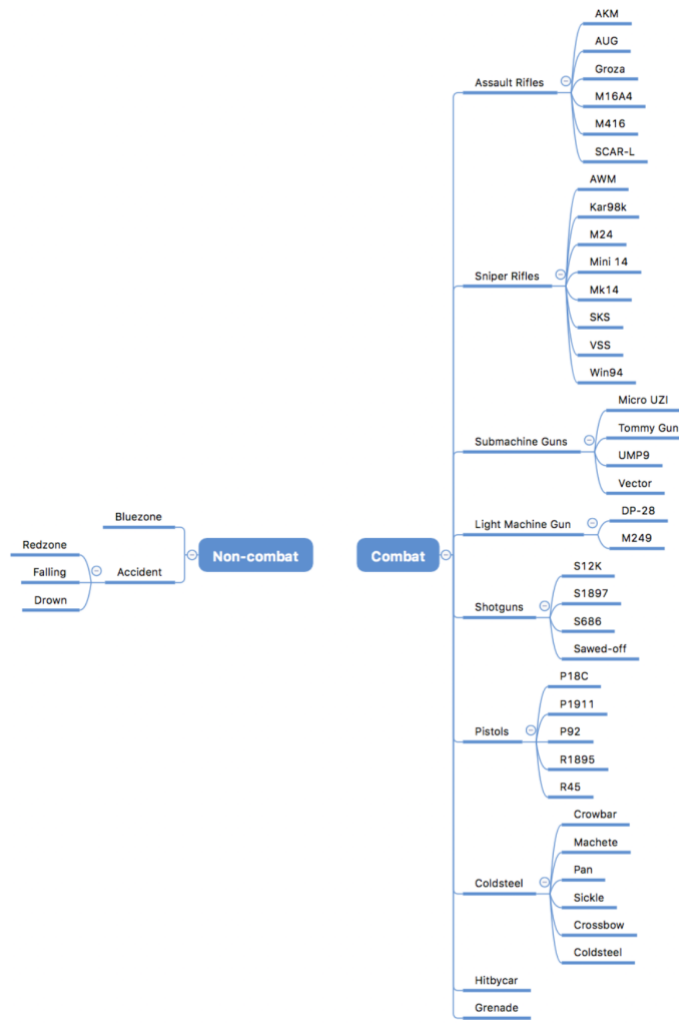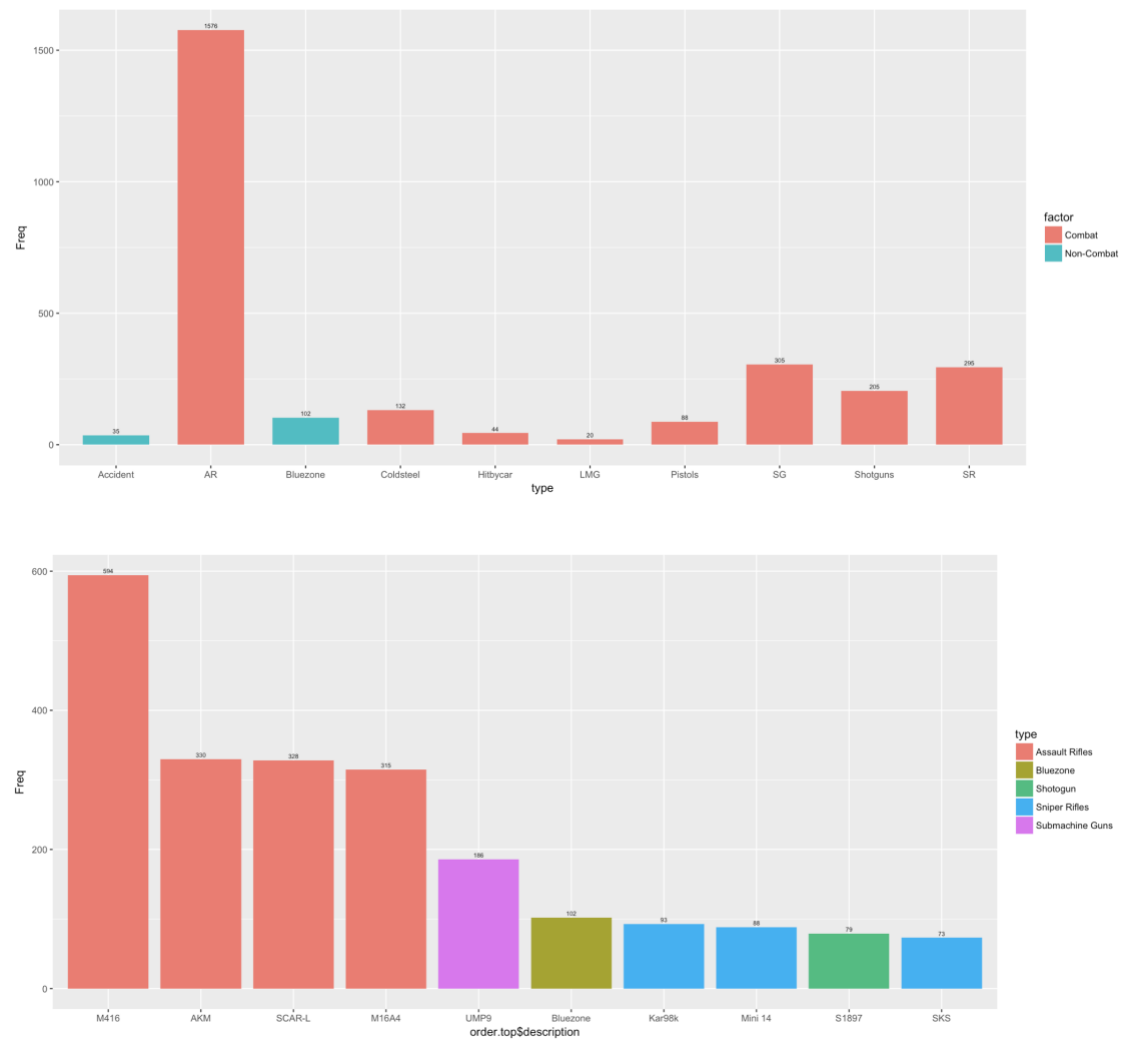
- Data Manipulation
  - ◆ Filter the deaths case by description='description '
  - ◆ Group together add extra attributes by factors
  - ◆ Sum up to calculate the frequency

- Visual Encoding

Plot figure as bar charts by groups or weapons, indicate the frequency by length of bar, categorise types by colours.
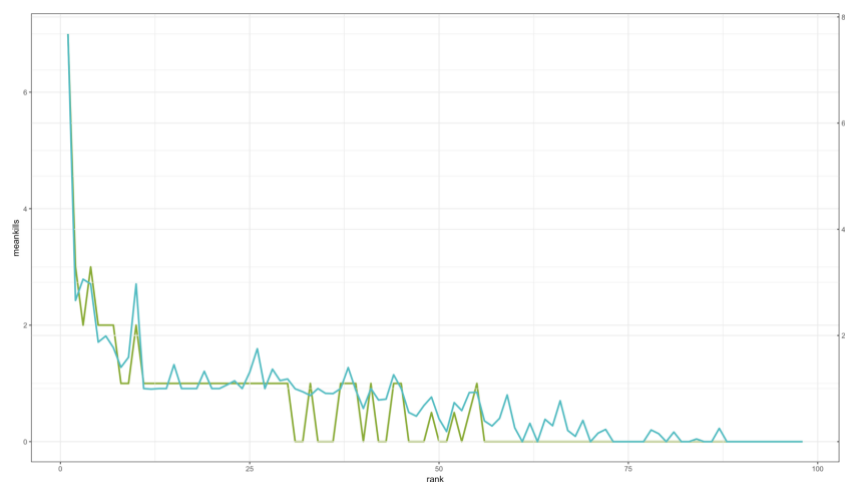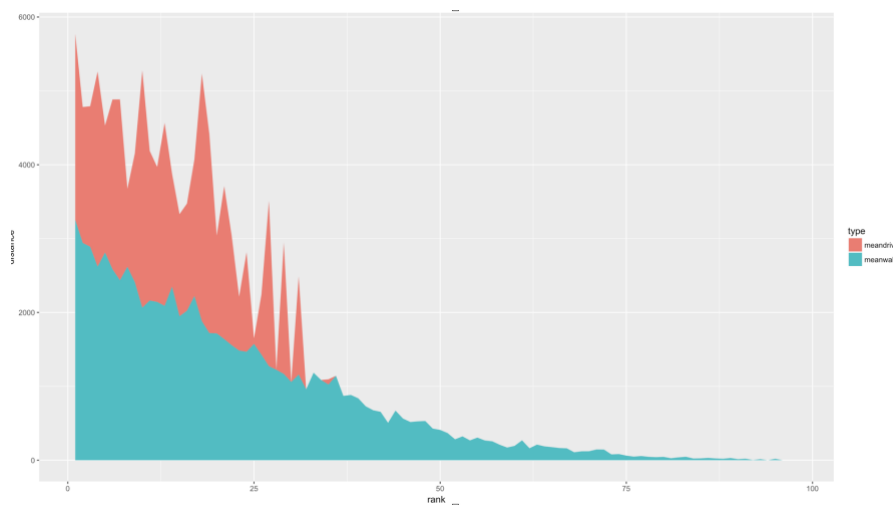
● Analysis and Answers

The first figure shows that Combat factors accounts for the most death while Assault Rifles is leading ahead. The second picture illustrate that M416 is the most popular weapon used by players and the number of other weapons.

Q3: What's the relationship between rank moving distance? What's the relationship between rank and kills/damages?

● Data Manipulation

◆ Select stats attributes, 'rank', 'walk_distance', 'ride_distance' and add new attribute as 'total_distance'.Sum cases up and get the mean of walk_distance and ride_distance by rank

◆ Select stats attributes, 'rank', 'damage_dealt' and add new attribute as 'total_distance'. Sum cases up and get the mean of damage_dealt and total_distance by rank.

● Visual Encoding

　　◆ Plot move distance as Area chart with rank as X axis, figure and types are indicated by the area and colour respectively.

　　◆ Plot figures as Line graph with double Y axis.

　　◆

● Analysis and Answers:

　　◆ The first graph shows the relationship between rank and the move distance is that the player who survive longer usually using vehicle to march, as approximately top 30 players' drive distance take up around 1/2 in total.

◆ The second plot illustrate that the similarity between damage and kills but not the same. But considering to the relationship with rank, only 10 %, top15 players, kills more opponents and the figure goes up dramatically to the winner.

Exploratory question:

The generate process of following question is related to the results of Q2&3, and the mechanism of game. The question is a deep looking into the former question.

Q4: Is damage in correlated to kills?

(Data manipulation& Visual encoding is same in Question 3)

Q5: What's the characteristics of different types weapon? How do they used in the combat?

● Data manipulation:
  ◆ Based on Q2, there is another related attribution added, which is 'distance', calculated by locations.
  ◆ Sum cases up and get the mean of time_event and distance by group.
  ◆ Selecti four major types of weapons (Assault Rifles, Sniper Rifles, Light Machine Gun, Shotguns), simply plot by every cases.
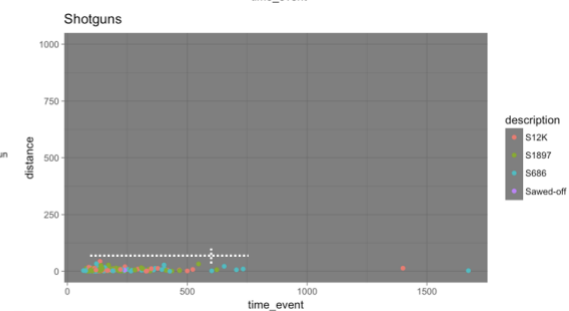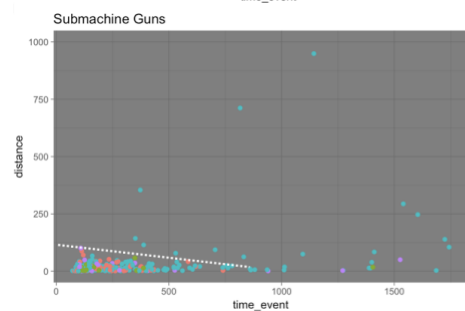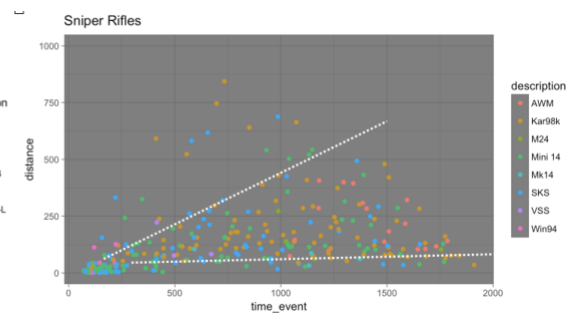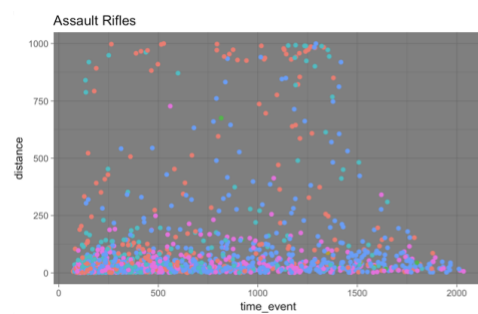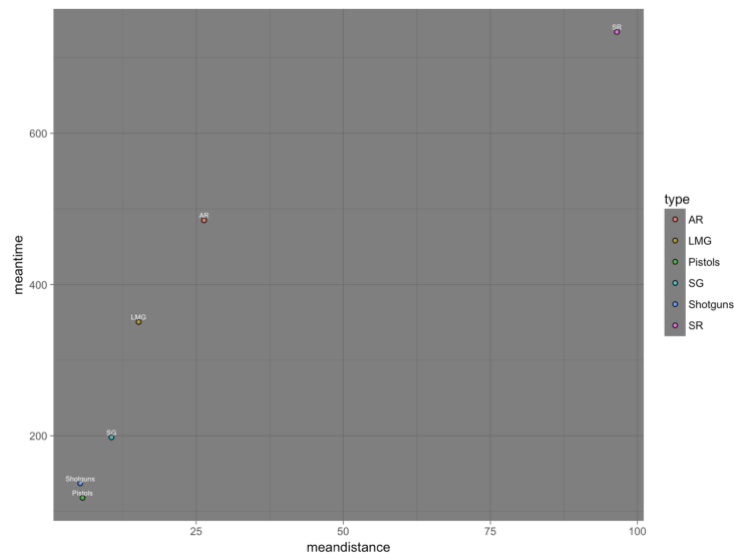
- Visual Encoding
  - ◆ 1.Plot the mean figure as scatter by types of group, indicate the figure by position, categorise group by colours.
  - ◆ 2.Plot every cases as scatter by detailed types of weapon, indicate the figure by position, categorise types by colours.

- Analysis and Answers
  - ◆ The first figure generally shows kinds of weapon's feature as the applied distance and used time.

◆ The further conclusion can been draw from the second picture, which is, for example, shotguns is mostly used in short distance and early stage while AR covers more scopes of combat. The Sniper rifles is applied more in long distance when game goes on, .etc.

Discussion of visualization Design

From the perspective of Question 1, the variable is about the position and time. Therefore, plot the scatters on the game map could be expressive and effective to perceive. Regarding time is a wide range interval variable, hue could be applied. However, when analyse the intensiveness of the battle field, heated may be a better way. But more works need to be done, like detailed dividing map place and grouping position

From the perspective of Question 2 and 5, the variable involves ordinal variable, which has several values. The statistic is to get the frequencies and mean value by dimensions. The visualisation method is to transfer the measurement into the position of the on the graph, which provides an intuitive way to compare. The points in the Question 5 plot are dense but the distribution of points is more specific to the answer the problem than the relative statistics. An ideal visualisation and could be zoomable.

From the perspective of Question 3 and 4, which are mainly about the visualization of the two variables, using line graphs or similar areas can be better visual positive or negative relationship. With the statistics, the plot obviously shows the same trend of the results and the two-coordinate plots can be well fitted. But further confirmation of the relationship between the two needs such as Pearson method.

Additionally, the size of dataset is considerable small, with only 31 matches and totally around 3000 obs. If there are more data, the method of data cleaning and visualization should be updated to fit.

Reflection on development process

- Data Crawling – Python

  The code of crawling program was not sufficient, since only collected 31match of the data. The process of crawling was focusing on one players' match record. These could cause potential bias.

- Manipulation - R

  In the early period of manipulating, the use of R is blunt. After getting familiar with plyr and dplyr, the manipulation is easier although the major part coding has been completed. As for the dimension, more relationship or interesting could be dived into if have enough data. For example, the observe point could be one players' overall performance in a series of matches given, if there is enough data

- Visualisation - ggplot

  gglot2 is an easy-to-use and powerful library which could meet visualization requirements in addition to the lack of interaction. During the development, the difficultis were the processing of the background image, to make it match the plot of the plot and merging the two line charts together.