



PUBG Firearms Analysis

Big Data and Cloud Computing

Final Project Report

Written by

Yangying Zhao, 489368

Baoxin Lai, 499100

Wenting Wang, 500867

Haidong Wang, 499996

Submitted to

Prof. HOSSEIN AMINI

December 10th, 2021

Executive Summary

In order to gain a competitive advantage in different distance battling, picking a right firearms combination is important in PUBG. This paper discusses the characteristics of popularity and killing distance of different guns through the division data. The dataset used is a proper fraction of PUBG Match Deaths and Statistics. First, for firearm frequent analysis, the report employed mapReduce to get the frequency of the 28 firearms, then the result was visualized by wordcloud and excel, as well as R for the classified firearms. Then for firearm kill distance analysis, the report first used mapReduce to achieve the maximum distance, minimum distance and average distance, and to dig deeper, the map function was applied to get all the distance for the firearms, then the results were visualized by excel and R for different firearm types as well.

With in-depth research, there are two firearms combinations recommendations considering long-term and all-distance battlegrounds. Firstly, Rifle + Submachine Gun/Shotgun. Secondly, Rifle + Sniper Rifle. In this game, all the weapons are randomly dropped on the ground, so picking weapons wisely is the key to win the game.

Description of the Data

PUBG(PlayerUnknown's Battlegrounds) is a first/third-person shooter battle royale style game that matches 100 players on a large map. Players parachute from an airplane onto towns and search buildings for weapons, ammo, armor and first-aid. Players will fight to be the last squad standing to win. (See Appendix A)

In the deaths file, this file records each death that occurred among 720k matchings. Each row represents each death in matching.

Variable Name	Meaning	Type
killed_by	The weapon killer used	Nominal
killer_name	Killer id in game	Nominal
killer_placement	The killer team's final ranking	Discrete
killer_position_x	Killer's location's x-coordinate when killing happen	Continuous
killer_position_y	Killer location's y-coordinate when killing happen	Continuous
map	Game map(either ERANGEL ISLAND or MIRAMAR)	Nominal
match_id	Each match unique ID	Nominal
time	The time last till the kill happens(count in seconds)	Discrete
victim_name	Victim id in game	Nominal
victim_placement	The Victim team's final ranking	Discrete
victim_position_x	Victim's location's x-coordinate when killing happen	Continuous
victim_position_y	Victim's location's y-coordinate when killing happen	Continuous

Problem Statement

1. Count the frequency of firearms used in all kills to see which kind or specific gun has a higher probability of killing the enemy.
2. Analyze the killing distances of different firearms types and combine each gun's characteristics to discuss which gun has a higher probability of killing the enemy under different battle distances.
3. Make recommendations on what firearms combination is better in increasing killing probability.

Why is this a big Data

The original PUBG Match Deaths and Statistics data set includes about 60 million observations and is about 20G in size, which is quite huge and cannot be handled by traditional analysis tools such as Excel. We extract 1/10 of the data where there are 13,426,349 death records.

Method & Results

1. Firearm Frequency Analysis

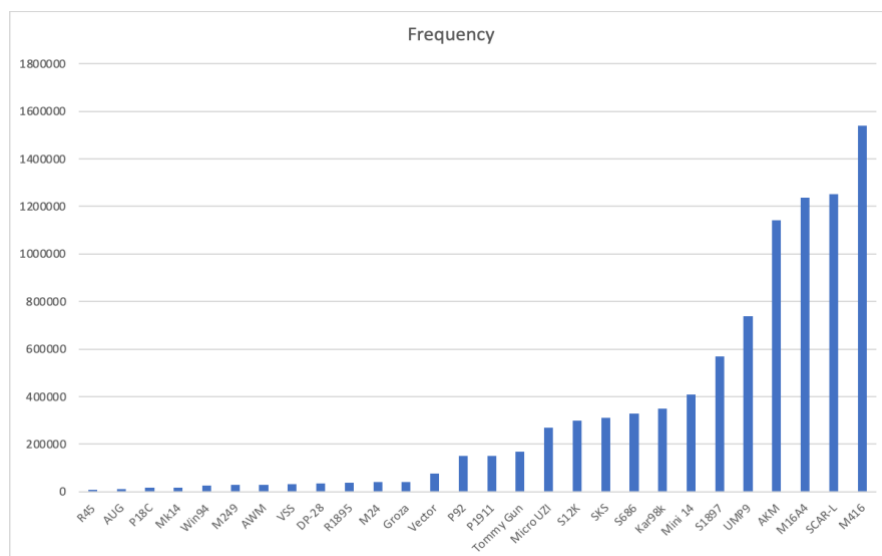
a. First use Map killmap.py to pick out all the lines whose “killed_by” belong to the firearm list and count 1 for every occurrence. Then use Reduce killred.py to count the frequency of each firearm.

(See Appendix B)

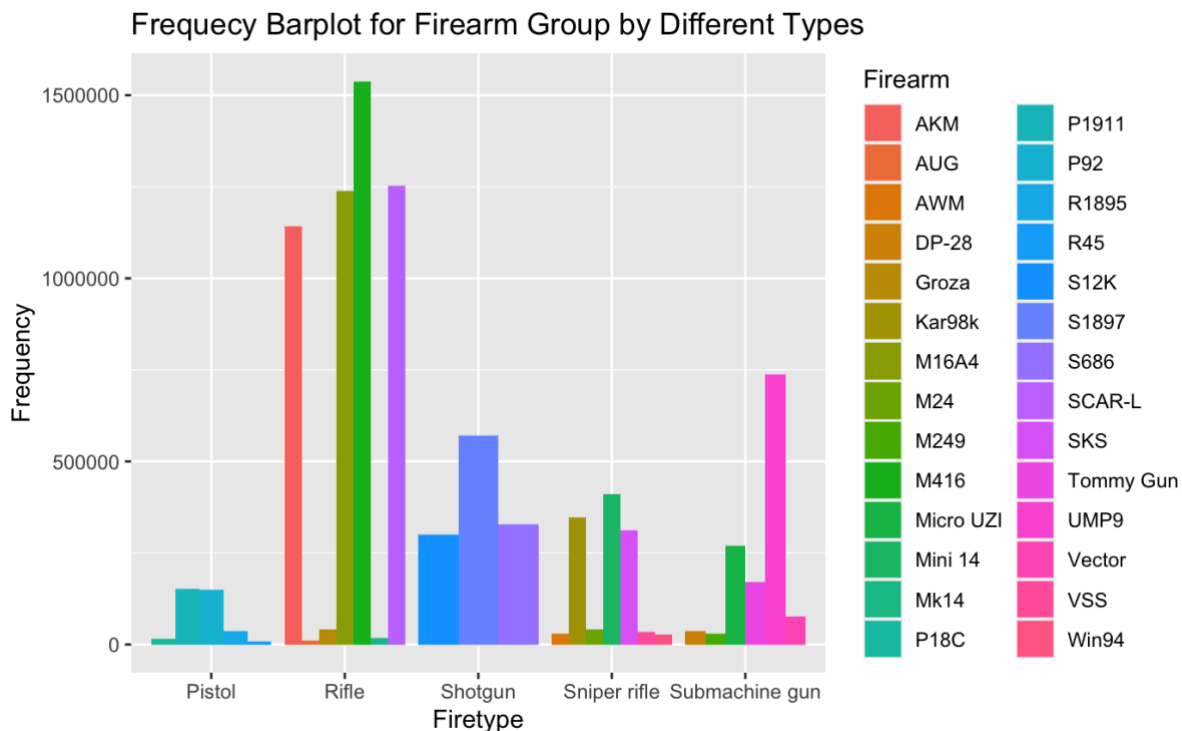
```
[yangying.z@ip-172-31-95-86 ~]$ hdfs dfs -cat firearm_freq/part-00000
SKS      311528
Micro UZI      268974
S686     329326
Mk14     16835
AKM      1142941
M249     29457
M16A4    1237998
P1911    151251
P92      149565
UMP9     737322
S12K     299421
AWM      29596
P18C     15952
DP-28    36182
AUG      11469
M24      41546
R1895    37779
Vector   77513
SCAR-L   1252632
Tommy Gun      169611
VSS      33170
Win94    26723
Mini 14   409859
M416     1538312
R45      8492
Kar98k   348114
Groza    42178
S1897    569420
```

b. Employ Excel to visualize the firearms’ frequency in ascending order.

Employ R to visualize the results while classifying the firearms into Sniper rifle, rifle, Submachine gun, Pistol and Shotgun. (See Appendix B)



Employ R to visualize the results while classifying the firearms into Sniper rifle, rifle, Submachine gun, Pistol and Shotgun. (See Appendix B)

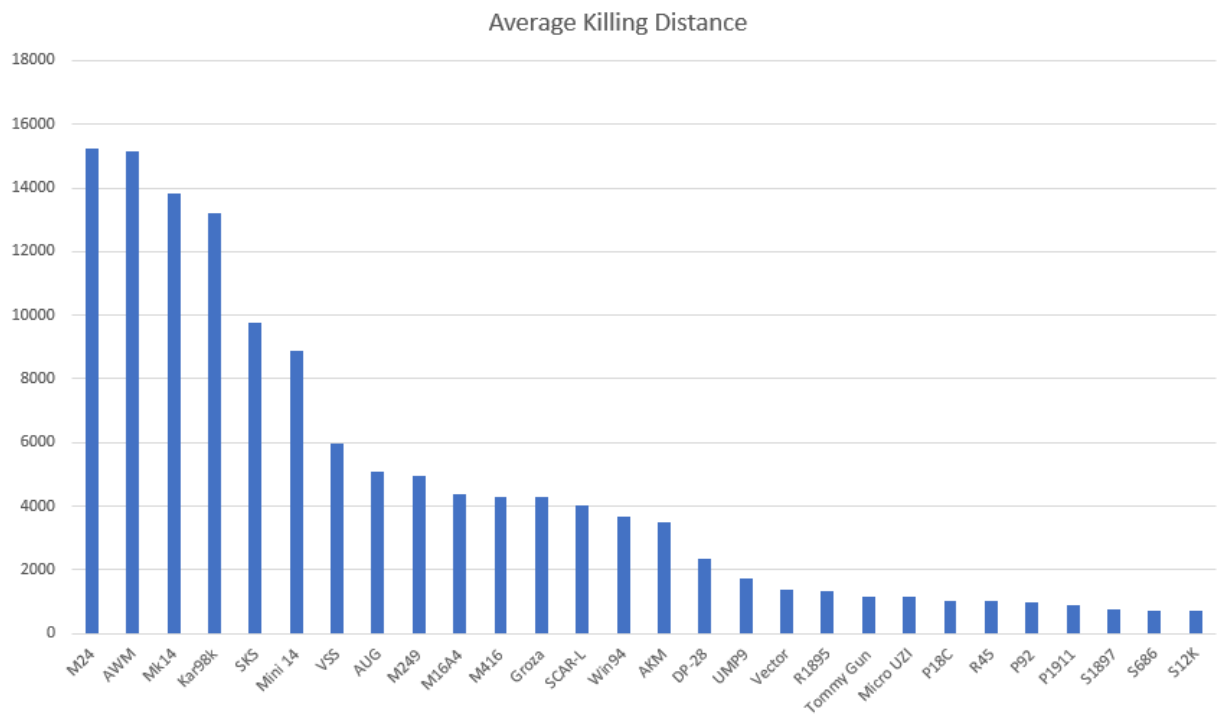


2. Firearm Kill Distance Analysis

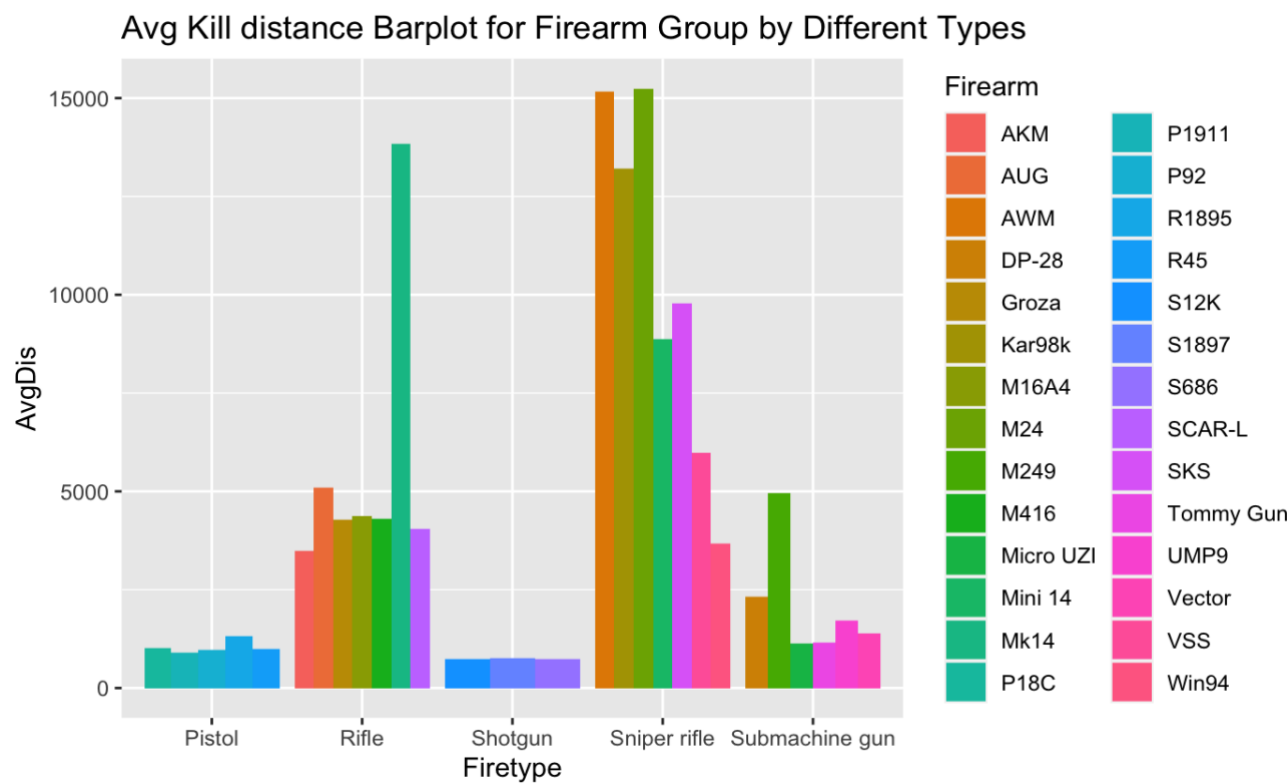
a.1 First use Map dismap.py to pick out all the lines whose “killed_by” belongs to the firearm list and calculate the distance with the function $\text{distance} = ((\text{killer_position_x} - \text{victim_position_x})^2 + (\text{killer_position_y} - \text{victim_position_y})^2)^{0.5}$. Then use Reduce disred.py to list the maximum distance, minimum distance and average distance for the 28 firearms. (See Appendix C)

```
[yangying.z@ip-172-31-95-86 ~]$ hdfs dfs -cat firearm_dis/part-00000
SKS 100064.321816 0.0 9777.59596486
Micro UZI 168602.732475 0.0 1133.4954793
S686 100297.203951 0.0 734.178124251
Mk14 99113.5070991 0.0 13832.480153
AKM 175000.314038 0.0 3493.8611881
M249 100144.964996 28.2708330263 4950.61645882
M16A4 189567.672822 0.0 4365.38841163
P1911 98334.9527434 0.0 905.144657493
P92 105084.939283 0.0 980.312376068
UMP9 105226.299025 0.0 1722.43046105
S12K 102010.682489 0.0 730.662194251
AWM 99683.3063575 62.5863403627 15164.8806978
P18C 99906.9614532 0.0 1024.93019675
DP-28 99307.4816228 0.0 2329.19782202
AUG 98630.5071286 0.0 5092.01492974
M24 100090.212814 0.0 15246.6868106
R1895 117900.70639 0.0 1310.36322251
Vector 100159.345032 0.0 1384.22297138
SCAR-L 103997.908923 0.0 4040.25957149
Tommy Gun 99558.1455731 0.0 1154.00070254
VSS 100319.071893 0.0 5975.9673523
Win94 99523.4334653 0.0 3684.91448318
Mini 14 115047.143056 0.0 8879.97989401
M416 179808.233113 0.0 4297.95095355
R45 17091.9955912 0.0 1000.81439942
Kar98k 101555.829116 0.0 13200.6709062
Groza 99317.7426241 0.0 4289.73543539
S1897 356309.873402 0.0 764.199962531
```

b.1 Employ Excel to visualize the firearms’ average distance in descending order.

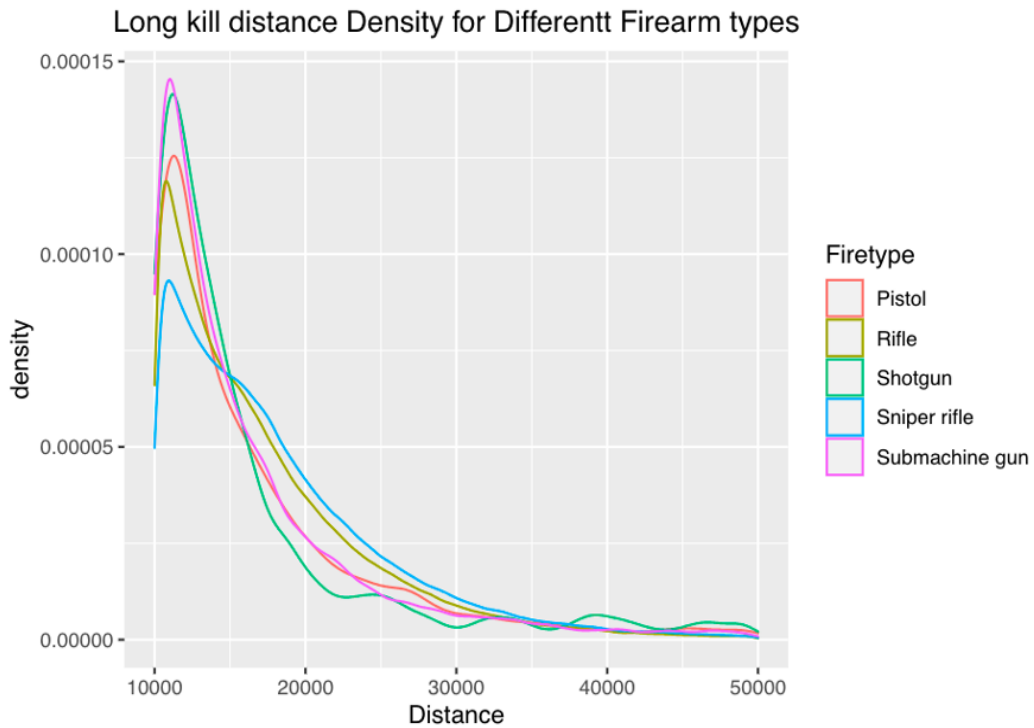
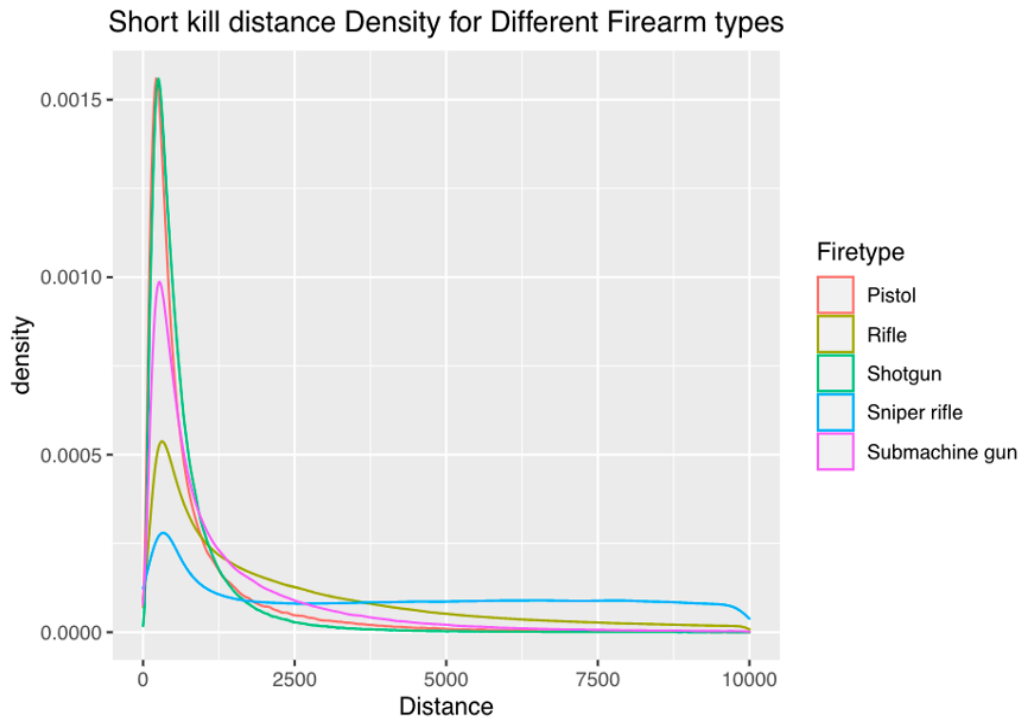


Employ R to visualize the results while classifying the firearms into Sniper rifle, rifle, Submachine gun, Pistol and Shotgun. (See Appendix C)



a.2 First use Map dismap1.py to pick out all the lines whose “killed_by” belong to the firearm list and calculate the distance with the function (same as above). Create a csv file firearm_dis.csv with the results and download the file from the server. (See Appendix D)

b.2 Employ R to visualize the results with boxplots and density maps, as well as classifying the firearms into Sniper rifle, rifle, Submachine gun, Pistol and Shotgun. (See Appendix D)



Conclusion

1. Firearms Using Frequency

After counting the frequency and sorting these 28 kinds of firearms, the ranking of guns used in killing the enemy is listed above. These 28 kinds of firearms and be divided into the following five types:

Sniper Rifles: M24, AWM, Kar98k, SKS, Mini 14, VSS, Win 94

Rifles: Mk14, AUG, M16A4, M416, Groza, SCAR-L, AKM

Submachine guns: M249, DP-28, UMP9, Vector, Tommy Gun, Micro UZI

Shotguns: S1897, S686, S12k

Pistols: R1895, P18C R45, P92, P1911

The frequency graph above shows that the top three guns used in killing enemies are all rifles. The following is the submachine gun. The sniper rifle and shotgun are nearly close to each other. The pistol is the last choice. The popularity of rifles may be because this is a big database that would count on all levels of players. Also, the rifle is comprehensive in all aspects, such as power, range, stability, and flexibility. The rifle has a high firing rate, long-range and large ammo capacity make the rifle have well-rounded performance. Due to its significant damage on both body and head shots, the top four ranking rifles mostly can make the two-shot kill. The rifle can do well with both close-distance combating and long-distance aiming. It is a must-have for almost all-level players, especially for beginners.

For the top-ranking weapon M416, the range and rate of fire are higher than non-equipped SCAR and lower than M16 a little bit. However, after will come with the butt, its stability and direct range are significantly more robust than other rifles after all the attachments. The high fault tolerance and performance of M416 made it the top-ranking killing weapon.

The submachine gun and shotguns' characteristic kills the enemy fastest at close range. All machinegun has almost no difference in stability, even though shotgun has higher power in damage. Still, a shotgun's shooting distance is short, which will talk about it in following shooting distance part, the distance of shotgun only half warehouse, compared to a submachine gun, it has 3 to 4 times of shooting distance of it. So the submachine gun has higher popularity than the shotgun.

The sniper rifle is great damage on the head shot. If shooting directly on the head most snipers can make a one-shot kill. However, this required players' shooting accuracy and this database player is from all levels, which might explain why sniper's popularity is low.

There is not much choice during the early stage, so the scenes that require a pistol are not very selective for players. The pistol is the least popular because after the other weapons are well equipped, the pistol is weak in power. The killing mostly happened during the middle and late stages, easily explaining that the pistol is the least popular firearm.

2. Firearms Kill Distance

The density and killing distances box plot for different types of firearms matches the real-world result perfectly, the fire distance ranking as follows: shotgun-pistol-submachine gun-rifle-sniper rifle. This means in the close range, the submachine gun and the shotgun are better for close combat; in the long-distance map, sniper rifles are more suitable for long-distance combat. If the player wants to stay alive longer till the final circle, a good combination of all distances is needed.

3. Firearms Combination Recommendation

Based on the previous analysis, there are two combination recommendations.

1) Rifle + Submachine Gun/ Shotgun

Despite the all distanced rifle, bringing a shotgun is really useful after landing in a place with a lot of people especially in a melee or defend in the building, a shotgun is the strongest weapon and allows you to kill with a single shot.

After the player gets close to the final circle, longer distance combat is required also closed distance needs greater power firearm, so changing from the shotgun to submachine gun can make better damage since the range of the submachine gun is about three to four times that of the shotguns, and it can adapt to more situations than the shotgun. It can be said that the spray is suitable for combat within 50M, and The best effect of the rifle lies in the mid-to-long range 200-400m battle, and the submachine gun's effect can be fully demonstrated in the 0-150m, and this distance is also a common distance indecisive battle circles or encounters.

2) Rifle + Sniper Rifle

Because the sniper rifle has the ability to kill with one shot at a long distance and strike at medium and long-range. The reason for this combination is that the rifle itself is not too bad in close combat. Generally speaking, if the player is a novice, you can't fully use the gun. All the characteristics, the gap between rifles and submachine guns is not very obvious, rifles also have the ability to kill at close range, so many people are willing to sacrifice a little melee advantage to obtain a medium and long-range strike capability. Whether it's electric is coming, or cleaning up the enemies on the path, it is very effective. Long-range sniper rifles have absolute dominance.

Appendix A

PUBG Match Deaths and Statistics, <https://www.kaggle.com/skihikingkevin/pubg-match-deaths>

Appendix B

nano killmap.py

```
GNU nano 2.3.1 File: killmap.py Modified
#!/usr/bin/env python

import sys

firearms = ['P18C','P1911','P92','R1895','R45','Win94','S1897','S686','S12K','UMP9','Micro UZI','Vector','Tommy Gun','AKM','M416','SCAR-L','M16A4','AUG',
            'Groza','M249','DP-28','VSS','Mini 14','SKS','Mk14','Kar98k','M24','AWM']

for line in sys.stdin:
    line = line.strip()
    killby = line.split(',')[0]
    if killby in firearms:
        print '%s\t%s' % (killby, 1)
```

nano killred.py

```
GNU nano 2.3.1 File: killred.py Modified
#!/usr/bin/env python

import sys

firearm_freq = {}
for line in sys.stdin:
    firearm, count = line.strip().split('\t')
    try:
        count = int(count)
    except ValueError:
        continue
    try:
        firearm_freq[firearm] += count
    except:
        firearm_freq[firearm] = count

for firearm in firearm_freq.keys():
    print '%s\t%s' % (firearm, firearm_freq[firearm])
```

nano killrunmr.sh

```
GNU nano 2.3.1 File: killrunmr.sh Modified
#!/bin/bash
hadoop jar /opt/cloudera/parcels/CDH-7.1.7-1.cdh7.1.7.p0.15945976/jars/hadoop-streaming-3.1.1.7.1.7.0-551.jar \
    -Dmapred.reduce.tasks=1 \
    -input /user/yangying.z/kill.csv \
    -output /user/yangying.z/firearm_freq \
    -file killmap.py \
    -file killred.py \
    -mapper "python killmap.py" \
    -reducer "python killred.py"
```

```
[yangying.z@ip-172-31-95-86 ~]$ chmod +x killmap.py
[yangying.z@ip-172-31-95-86 ~]$ chmod +x killred.py
[yangying.z@ip-172-31-95-86 ~]$ nano killrunmr.sh
[yangying.z@ip-172-31-95-86 ~]$ hdfs dfs -put killmap.py
[yangying.z@ip-172-31-95-86 ~]$ hdfs dfs -put killred.py
[yangying.z@ip-172-31-95-86 ~]$ hdfs dfs -put kill.csv
[yangying.z@ip-172-31-95-86 ~]$ hdfs dfs -put killrunmr.sh
[yangying.z@ip-172-31-95-86 ~]$ bash killrunmr.sh
```

bash killrunmr.sh

```
[yangying.z@ip-172-31-95-86 ~]$ hdfs dfs -ls firearm_freq
Found 2 items
-rw-r--r--  3 yangying.z yangying.z      0 2021-12-07 05:42 firearm_freq/_SUCCESS
-rw-r--r--  3 yangying.z yangying.z    344 2021-12-07 05:42 firearm_freq/part-00000
```

hdfs dfs -cat firearm_freq/part-00000

```
[yangying.z@ip-172-31-95-86 ~]$ hdfs dfs -cat firearm_freq/part-00000
SKS      311528
Micro UZI      268974
S686     329326
Mk14     16835
AKM      1142941
M249     29457
M16A4    1237998
P1911    151251
P92      149565
UMP9     737322
S12K     299421
AWM       29596
P18C     15952
DP-28    36182
AUG       11469
M24       41546
R1895    37779
Vector   77513
SCAR-L   1252632
Tommy Gun 169611
VSS       33170
Win94     26723
Mini 14  409859
M416     1538312
R45       8492
Kar98k   348114
Groza    42178
S1897    569420
```

R code for visualization

```
library(readxl)
```

```
freq = read_excel("Firearm_freq.xlsx")
```

add a column to classify the firearms

```
freq$Firetype[freq$Firearm %in% c('Kar98k','M24','AWM','SKS','Mini 14','VSS','Win94')] =  
'Sniper rifle'
```

```
freq$Firetype[freq$Firearm %in% c('Mk14','AUG','M16A4','Groza','SCAR-L','AKM','M416')] =  
'Rifle'
```

```
freq$Firetype[freq$Firearm %in% c('M249','DP-28','UMP9','Vector','Tommy Gun','Micro UZI')] =  
'Submachine gun'
```

```
freq$Firetype[freq$Firearm %in% c('R1895','P18C','R45','P92','P1911')] = 'Pistol'
```

```
freq$Firetype[freq$Firearm %in% c('S1897','S686','S12K')] = 'Shotgun'
```

plot the frequency

```
library(ggplot2)
```

```
library(ggthemes)
```

use barplot to view the general difference of frequency among different firearm types

```
ggplot(data = freq, aes(Firetype, Frequency, fill = Firearm)) + geom_bar(position = "dodge", stat =
"identity") + labs(title = 'Frequency Barplot for Firearm Group by Different Types')
```

Appendix C

nano dismap.py

```
GNU nano 2.3.1 File: dismap.py Modified

#!/usr/bin/env python

import sys

firearms = ['P18C', 'P1911', 'P92', 'R1895', 'R45', 'Win94', 'S1897', 'S686', 'S12K', 'UMP9', 'Micro UZI', 'Vector', 'Tommy Gun', 'AKM', 'M$
Groza', 'M249', 'DP-28', 'VSS', 'Mini 14', 'SKS', 'Mk14', 'Kar98k', 'M24', 'AWM']

for line in sys.stdin:
    line = line.strip()
    killby = line.split(',')[0]
    if killby in firearms:
        try:
            kill_x = float(line.split(',')[3])
        except ValueError:
            continue

        try:
            kill_y = float(line.split(',')[4])
        except ValueError:
            continue

        try:
            vic_x = float(line.split(',')[10])
        except ValueError:
            continue

        try:
            vic_y = float(line.split(',')[11])
        except ValueError:
            continue

        if kill_x != 0 and kill_y != 0 and vic_x != 0 and vic_y != 0:
            distance = ((kill_x - vic_x)**2 + (kill_y - vic_y)**2)**0.5
            print "%s\t%s" % (killby, float(distance))
```

nano disred.py

```
GNU nano 2.3.1 File: disred.py Modified

#!/usr/bin/env python

import sys

firearm_dis = {}
for line in sys.stdin:
    firearm, distance = line.strip().split("\t")
    try:
        dis = float(distance)
    except ValueError:
        continue
    try:
        firearm_dis[firearm].append(dis)
    except:
        firearm_dis[firearm] = [dis]

for firearm in firearm_dis.keys():
    max_dis = max(firearm_dis[firearm])
    min_dis = min(firearm_dis[firearm])
    avg_dis = sum(firearm_dis[firearm])/len(firearm_dis[firearm])
    print '%s\t%s\t%s\t%s' % (firearm, max_dis, min_dis, avg_dis)
```

nano disrunmr.sh

```

GNU nano 2.3.1                               File: disrunmr.sh                               Modified
#!/bin/bash
hadoop jar /opt/cloudera/parcels/CDH-7.1.7-1.cdh7.1.7.p0.15945976/jars/hadoop-streaming-3.1.1.7.1.7.0-551.jar \
  -Dmapred.reduce.tasks=1 \
  -input /user/yangying.z/kill.csv \
  -output /user/yangying.z/firearm_dis \
  -file dismap.py \
  -file disred.py \
  -mapper "python dismap.py" \
  -reducer "python disred.py"

```

```

[yangying.z@ip-172-31-95-86 ~]$ chmod +x dismap.py
[yangying.z@ip-172-31-95-86 ~]$ chmod +x disred.py
[yangying.z@ip-172-31-95-86 ~]$ hdfs dfs -put dismap.py
[yangying.z@ip-172-31-95-86 ~]$ hdfs dfs -put disred.py
[yangying.z@ip-172-31-95-86 ~]$ hdfs dfs -put disrunmr.sh
[yangying.z@ip-172-31-95-86 ~]$ bash disrunmr.sh

```

bash disrunmr.sh

hdfs dfs -ls firearm_dis

```

[yangying.z@ip-172-31-95-86 ~]$ hdfs dfs -ls firearm_dis
Found 2 items
-rw-r--r--   3 yangying.z yangying.z           0 2021-12-07 07:28 firearm_dis/_SUCCESS
-rw-r--r--   3 yangying.z yangying.z       1057 2021-12-07 07:28 firearm_dis/part-000000

```

hdfs dfs -cat firearm_dis/part-000000

```

[yangying.z@ip-172-31-95-86 ~]$ hdfs dfs -cat firearm_dis/part-000000
SKS      100064.321816   0.0    9777.59596486
Micro UZI    168602.732475   0.0    1133.4954793
S686      100297.203951   0.0    734.178124251
Mk14      99113.5070991   0.0    13832.480153
AKM       175000.314038   0.0    3493.8611881
M249      100144.964996   28.2708330263  4950.61645882
M16A4     189567.672822   0.0    4365.38841163
P1911     98334.9527434   0.0    905.144657493
P92       105084.939283   0.0    980.312376068
UMP9      105226.299025   0.0    1722.43046105
S12K      102010.682489   0.0    730.662194251
AWM       99683.3063575   62.5863403627  15164.8806978
P18C      99906.9614532   0.0    1024.93019675
DP-28     99307.4816228   0.0    2329.19782202
AUG       98630.5071286   0.0    5092.01492974
M24       100090.212814   0.0    15246.6868106
R1895     117900.70639    0.0    1310.36322251
Vector    100159.345032   0.0    1384.22297138
SCAR-L    103997.908923   0.0    4040.25957149
Tommy Gun  99558.1455731   0.0    1154.00070254
VSS       100319.071893   0.0    5975.9673523
Win94     99523.4334653   0.0    3684.91448318
Mini 14   115047.143056   0.0    8879.97989401
M416      179808.233113   0.0    4297.95095355
R45       17091.9955912   0.0    1000.81439942
Kar98k    101555.829116   0.0    13200.6709062
Groza     99317.7426241   0.0    4289.73543539
S1897     356309.873402   0.0    764.199962531

```

R code for visualization

```
avgdis = read_excel("Firearm_avgdis.xlsx")[, c(1,4)]
```

```
colnames(avgdis)[2] = 'AvgDis'
```

add a column to classify the firearms

```
avgdis$Firetype[avgdis$Firearm %in% c('Kar98k','M24','AWM','SKS','Mini 14','VSS','Win94')] =  
'Sniper rifle'
```

```
avgdis$Firetype[avgdis$Firearm %in% c('Mk14','AUG','M16A4','Groza','SCAR-L','AKM','M416')]  
= 'Rifle'
```

```
avgdis$Firetype[avgdis$Firearm %in% c('M249','DP-28','UMP9','Vector','Tommy Gun','Micro  
UZI')] = 'Submachine gun'
```

```
avgdis$Firetype[avgdis$Firearm %in% c('R1895','P18C','R45','P92','P1911')] = 'Pistol'
```

```
avgdis$Firetype[avgdis$Firearm %in% c('S1897','S686','S12K')] = 'Shotgun'
```

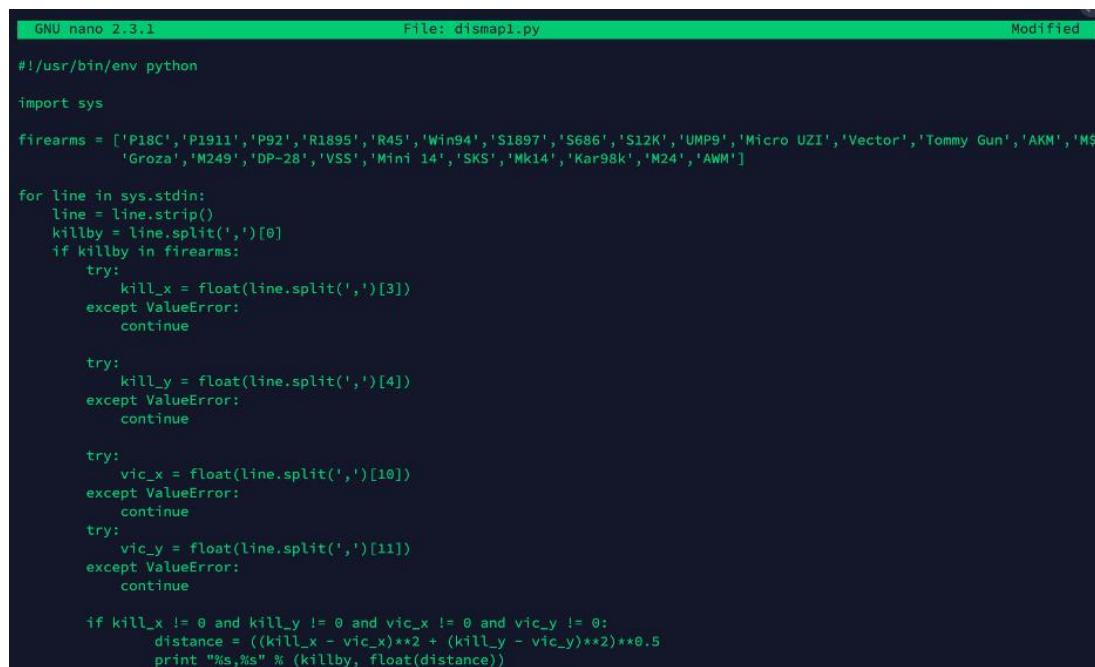
use barplot to view the general difference of average kill distance among different firearm types

```
ggplot(data = avgdis, aes(Firetype, AvgDis, fill = Firearm)) + geom_bar(position = "dodge", stat =  
"identity") + labs(title = 'Avg Kill distance Barplot for Firearm Group by Different Types')
```

```
View(avgdis)
```

Appendix D

nano dismap1.py



```
GNU nano 2.3.1 File: dismap1.py Modified  
#!/usr/bin/env python  
  
import sys  
  
firearms = ['P18C','P1911','P92','R1895','R45','Win94','S1897','S686','S12K','UMP9','Micro UZI','Vector','Tommy Gun','AKM','M416','Groza','M249','DP-28','VSS','Mini 14','SKS','Mk14','Kar98k','M24','AWM']  
  
for line in sys.stdin:  
    line = line.strip()  
    killby = line.split(',')[0]  
    if killby in firearms:  
        try:  
            kill_x = float(line.split(',')[3])  
        except ValueError:  
            continue  
  
        try:  
            kill_y = float(line.split(',')[4])  
        except ValueError:  
            continue  
  
        try:  
            vic_x = float(line.split(',')[10])  
        except ValueError:  
            continue  
  
        try:  
            vic_y = float(line.split(',')[11])  
        except ValueError:  
            continue  
  
    if kill_x != 0 and kill_y != 0 and vic_x != 0 and vic_y != 0:  
        distance = ((kill_x - vic_x)**2 + (kill_y - vic_y)**2)**0.5  
        print "%s,%s" % (killby, float(distance))
```



```
[yangying.z@ip-172-31-95-86 ~]$ cat kill.csv | python dismap1.py | head
SCAR-L,1422.5744009
S686,559.523413272
M416,243.625491277
AKM,21032.2831462
S686,68.20183282
S686,471.957635811
AKM,936.845905152
P92,419.421804393
S686,296.904311185
M416,151.01602564
```

cat kill.csv | python dismap1.py > firearm_dis.csv

```
[yangying.z@ip-172-31-95-86 ~]$ cat kill.csv | python dismap1.py > firearm_dis.csv
```

scp -i S_keypair.pem yangying.z@18.206.158.228:firearm_dis.csv ~/Desktop

```
(base) Yangyings-MacBook-Pro:Downloads yunazhaoyaya$ scp -i S_keypair.pem yangying.z@18.206.158.228:firearm_dis.csv ~/Desktop
firearm_dis.csv 100% 171MB 3.1MB/s 00:55
```

R code for Visualization

import and view data

```
dis = read.csv("firearm_dis.csv", header = FALSE)
```

```
colnames(dis) = c("Firearm", "Distance")
```

View(dis)

add a column to classify the firearms

```
dis$Firetype[dis$Firearm %in% c('Kar98k','M24','AWM','SKS','Mini 14','VSS','Win94')] = 'Sniper rifle'
```

```
dis$Firetype[dis$Firearm %in% c('Mk14','AUG','M16A4','Groza','SCAR-L','AKM','M416')] = 'Rifle'
```

```
dis$Firetype[dis$Firearm %in% c('M249','DP-28','UMP9','Vector','Tommy Gun','Micro UZI')] = 'Submachine gun'
```

```
dis$Firetype[dis$Firearm %in% c('R1895','P18C','R45','P92','P1911')] = 'Pistol'
```

```
dis$Firetype[dis$Firearm %in% c('S1897','S686','S12K')] = 'Shotgun'
```

plot the distance

```
library(ggplot2)
```

```
library(ggthemes)
```

use boxplot to view the general difference of distance among different firearm types

```
ggplot(data = dis, aes(Firetype, Distance, fill = Firearm)) +
```

```
  geom_boxplot(outlier.colour = 'grey') + coord_cartesian(ylim = c(0, 7*10^4)) +
```



```

labs(title = 'Kill distance Boxplot for Firearm Group by Different Types')

# use density map to look at the density of distance for different firearm types

# divide the range of distance into 0~10000 and 10000~50000

ggplot(data = dis, aes(x = Distance, colour = Firetype)) +

  geom_density() + scale_x_continuous(limits = c(0,10000)) +

  labs(x = 'Distance') + labs(title = 'Short kill distance Density for Different Firearm types') +
  theme(plot.title = element_text(hjust = 0.5))

ggplot(data = dis, aes(x = Distance, colour = Firetype)) +

  geom_density() + scale_x_continuous(limits = c(10000, 50000)) +

  labs(x = 'Distance') + labs(title = 'Long kill distance Density for Differentt Firearm types') +
  theme(plot.title = element_text(hjust = 0.5))

```