Problem Statement

Implement the use case present in below blog link

https://acadgild.com/blog/pig-use-case-pokemon-data-analysis/

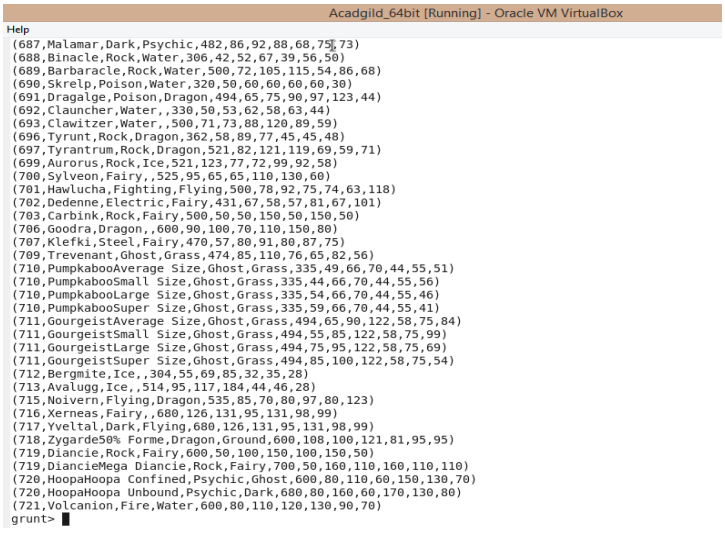
Problem1:

Find the list of players that have been selected in the qualifying round (DEFENCE>55).

Solution:

```
--Load Pokemon.csv Data
pokemon_data = LOAD '/home/acadgild/ankita/Assignment5_3/Pokemon' USING PigStorage(',')
AS(Sno:int,Name:chararray,Type1:chararray,Type2:chararray,Total:int,HP:int,Attack:int,Defense:int,splatck:int
,spldefenct:int,speed:int);

-- filter qualifying data
qualifying_data = FILTER pokemon_data by Defense>55 ;

-- display results

dump qualifying_data;
```

Output:



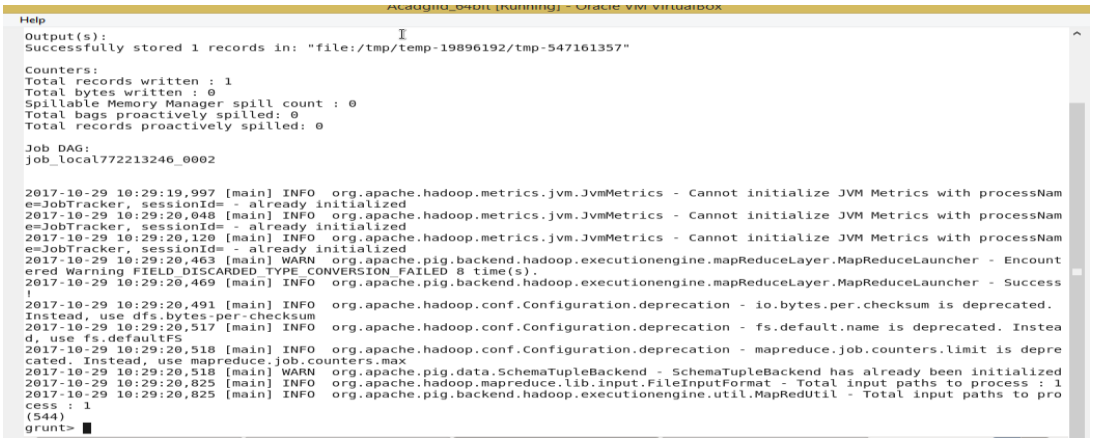
Problem2:

State the number of players taking part in the competition after getting selected in the qualifying round.

Solution:

```
-- Load Data
pokemon_data = LOAD '/home/acadgild/ankita/Assignment5_3/Pokemon' USING PigStorage(',')
AS(Sno:int,Name:chararray,Type1:chararray,Type2:chararray,Total:int,HP:int,Attack:int,Defense:int,splatck:int
,spldefenct:int,speed:int);

-- filter qualifying data
qualifying_data = FILTER pokemon_data by Defense>55 ;

-- group the qualifying data
group_qualifying_data = GROUP qualifying_data ALL ;

-- get count
count_qualifying_data = foreach group_qualifying_data GENERATE COUNT(qualifying_data) ;

-- display results
dump count_qualifying_data ;
```

Output:



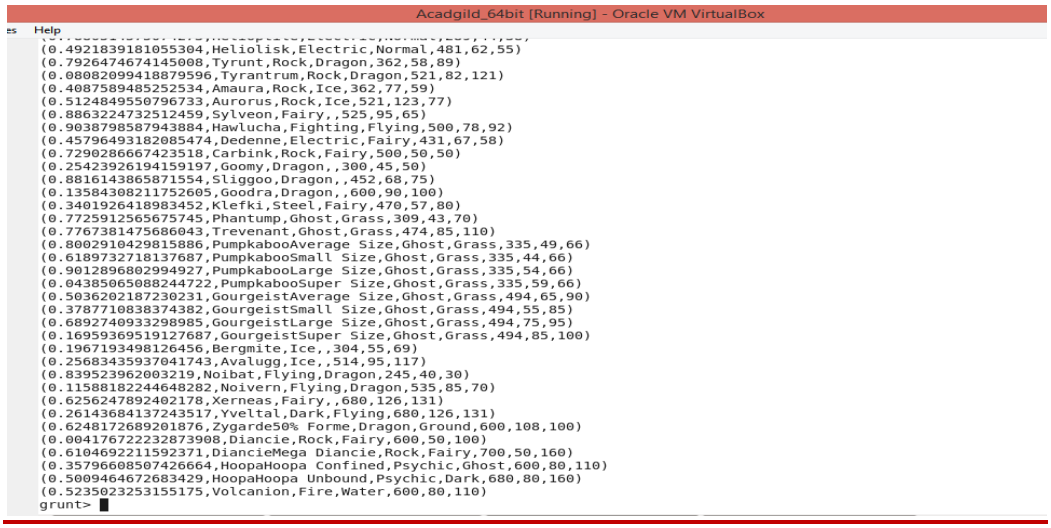
Problem3:

Using random() generate random numbers for each Pokémon on the selected list.

Solution:

```
-- Load Data
pokemon_data = LOAD '/home/acadgild/ankita/Assignment5_3/Pokemon'  USING PigStorage(',')
AS(Sno:int,Name:chararray,Type1:chararray,Type2:chararray,Total:int,HP:int,Attack:int,Defense:int,splatck:int
,spldefence:int,speed:int);

-- filter qualifying data

qualifying_data = FILTER pokemon_data by Defense>55 ;

-- generate random for relation 1
random_include1 = foreach qualifying_data GENERATE
RANDOM(),Name,Type1,Type2,Total,HP,Attack,Defense,splatck,spldefence,speed;

-- display result for random relation 1
dump random_include1;
```

Output:



Problem4:

Arrange the new list in a descending order according to a column randomly.

Solution:

```
-- Load Data
pokemon_data = LOAD '/home/acadgild/ankita/Assignment5_3/Pokemon'  USING PigStorage(',')
AS(Sno:int,Name:chararray,Type1:chararray,Type2:chararray,Total:int,HP:int,Attack:int,Defense:int,splatck:int
,spldefence:int,speed:int);

--Filter qualifying data

qualifying_data = FILTER pokemon_data by Defense>55 ;

-- generate random for relation 1
random_include1 = foreach qualifying_data GENERATE
RANDOM(),Name,Type1,Type2,Total,HP,Attack,Defense,splatck,spldefence,speed;

-- sort by random function relation 1
random1_desending = ORDER random_include1 BY $0 DESC;

-- displaying result of random function relation 1
dump random1_desending;
```

Output:

```
(0.042942226537938 11,Spheal,Ice,Water,290,70,40)
(0.04165872846874874,Amoonguss,Grass,Poison,464,114,85)
(0.04112818967833243,Liepard,Dark,,446,64,88)
(0.04036142442681656,Rhydon,Ground,Rock,485,105,130)
(0.040173611171296475,Magnemite,Electric,Steel,325,25,35)
(0.039372097531182204,Dusknoir,Ghost,,525,45,100)
(0.037787099482793685,Sunflora,Grass,,425,75,75)
(0.037366973223345146,Swablu,Normal,Flying,310,45,40)
(0.03729334521797689,Pyroar,Fire,Normal,507,86,68)
(0.036455393413030945,Numel,Fire,Ground,305,60,60)
(0.035944131069604035,ManectricMega Manectric,Electric,,575,70,75)
(0.03493765331300447,Roselia,Grass,Poison,400,50,60)
(0.03160041474391129,Simisage,Grass,,498,75,98)
(0.029942553769914615,Xatu,Psychic,Flying,470,65,75)
(0.027883790227097305,Typhlosion,Fire,,534,78,84)
(0.027694738598617152,Stunky,Poison,Dark,329,63,63)
(0.025001917548661323,MeloettaPirouette Forme,Normal,Fighting,600,100,128)
(0.024652717996032414,Bastiodon,Rock,Steel,495,60,52)
(0.022148741829219776,Fennekin,Fire,,307,40,45)
(0.021782870439575 72,Frillish,Water,Ghost,335,55,40)
(0.021612787914864873,Makuhita,Fighting,,237,72,60)
(0.019561773562484674,Heliolisk,Electric,Normal,481,62,55)
(0.018044909812275356,Sentret,Normal,,215,35,46)
(0.014442984279749105,Blissey,Normal,,540,255,10)
(0.013950454091826159,Scolipede,Bug,Poison,485,60,100)
(0.012985361702545717,Charmeleon,Fire,,405,58,64)
(0.009944563409446605,Archeops,Rock,Flying,567,75,140)
(0.009788104510038975,Lumineon,Water,,460,69,69)
(0.009692108790563925,Beedrill,Bug,Poison,395,65,90)
(0.007265466103542506,Dustox,Bug,Poison,385,60,50)
(0.004670869161197633,Deerling,Normal,Grass,335,60,60)
(0.0024079473065039547,MeowsticFemale,Psychic,,466,74,48)
grunt>
```

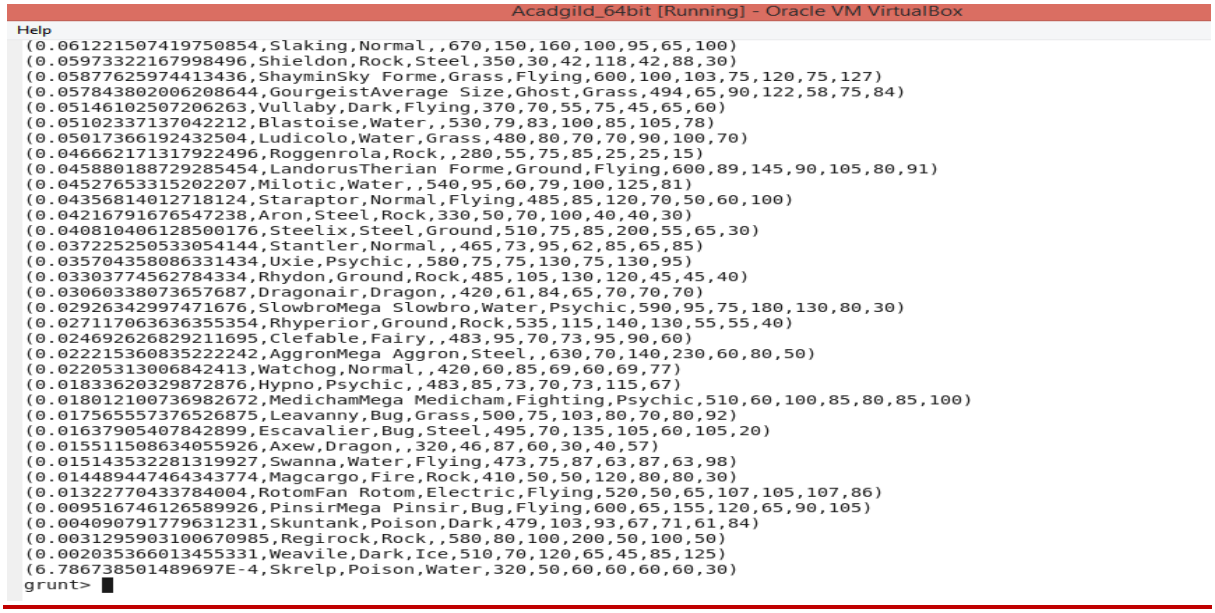
Problem5:

Now on a new relation again associate random numbers for each Pokémon and arrange in descending order according to column random.

Solution:

```
-- Load Data
pokemon_data = LOAD '/home/acadgild/ankita/Assignment5_3/Pokemon'  USING PigStorage(',')
AS(Sno:int,Name:chararray,Type1:chararray,Type2:chararray,Total:int,HP:int,Attack:int,Defense:int,splatck:int
,spldefence:int,speed:int);

qualifying_data = FILTER pokemon_data by Defense>55 ;

-- generate random for relation 2
random_include2 = foreach qualifying_data GENERATE
RANDOM(),Name,Type1,Type2,Total,HP,Attack,Defense,splatck,spldefence,speed;

-- sort by random function relation 2
random2_desending = ORDER random_include2 BY $0 DESC;

-- display results of  random function  for relation 2
dump random2_desending;
```

Output:



Problem6:

From the two different descending lists of random Pokémons, select the top 5 Pokémons for 2 different players.

Solution:

```
--Limit random function relation1

limit_random1 = LIMIT random1_desending 5;


--display result for random function for relation1

dump limit_random1;

--Limit random function relation2

limit_random2 = LIMIT random1_desending 5;


--display result for random function for relation1

dump limit_random2;
```

Output:

Result for random relation1---

```
2017-10-29 12:17:34,345 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Tota
2017-10-29 12:17:34,346 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUti
cess : 1
(0.9996741787224201,Milotic,Water,,540,95,60,79,100,125,81)
(0.9988937827689965,LucarioMega Lucario,Fighting,Steel,625,70,145,88,140,70,112)
(0.9909755224168616,Gligar,Ground,Flying,430,65,75,105,35,65,85)
(0.9902450737069991,Empoleon,Water,Steel,530,84,86,88,111,101,60)
(0.9854132973103437,Gorebyss,Water,,485,55,84,105,114,75,52)
grunt>
```

Result for random relation2---

```
2017-10-29 12:24:31,898 [main] INFO  org.apache.pig.backend.hadoop.executionengine.
cess : 1
(0.9994410696471268,Genesect,Bug,Steel,600,71,120,95,120,95,99)
(0.9974946918209349,Kyurem,Dragon,Ice,660,125,130,90,130,90,95)
(0.9974218714984913,Dedenne,Electric,Fairy,431,67,58,57,81,67,101)
(0.9967180869498755,Mesprit,Psychic,,580,80,105,105,105,105,80)
(0.9901542597845593,Torterra,Grass,Ground,525,95,109,105,75,85,56)
grunt>
```

Problem7:

Store the data on a local drive to announce for the final match. By the name player1 and player2 (only show the NAME and HP).

Solution:

```
-- Load Data
pokemon_data = LOAD '/home/acadgild/ankita/Assignment5_3/Pokemon'  USING PigStorage(',')
AS(Sno:int,Name:chararray,Type1:chararray,Type2:chararray,Total:int,HP:int,Attack:int,Defense:int,splatck:int
,spldefence:int,speed:int);

qualifying_data = FILTER pokemon_data by Defense>55 ;

-- generate random for relation 1
random_include1 = foreach qualifying_data GENERATE
RANDOM(),Name,Type1,Type2,Total,HP,Attack,Defense,splatck,spldefence,speed;

-- sort by random function relation 1
random1_desending = ORDER random_include1 BY $0 DESC;

-- generate random for relation 2
random_include2 = foreach qualifying_data GENERATE
RANDOM(),Name,Type1,Type2,Total,HP,Attack,Defense,splatck,spldefence,speed;

-- sort by random function relation 2
random2_desending = ORDER random_include2 BY $0 DESC;

-- top 5 from relation 1
Limit_random1 = LIMIT random1_desending 5 ;

-- top 5 from relation 2
limit_random2 = LIMIT random2_desending 5 ;

-- filter name from relation 1
filter_only_name1 = foreach limit_random1 Generate ($1,HP);

-- filter name from relation 2
filter_only_name2 = foreach limit_random2 Generate ($1,HP);

-- store relation 1 data to local file
STORE limit_random1 INTO '/home/acadgild/ankita/Assignment5_3/player1.txt';

-- store relation 2 data to local file
STORE limit_random2 INTO '/home/acadgild/ankita/Assignment5_3/player2.txt';
```

Output:

Result of Player1.txt—

```
!
grunt> cat Player1.txt;
0.9992667743956373      Malamar Dark    Psychic 482     86      92      88      68      75      73
0.9955287665740217      GlalieMega Glalie       Ice             580     80      120     80      120     80      100
0.9942589930765201      Charmeleon      Fire            405     58      64      58      80      65      80
0.9937828016672765      Gardevoir       Psychic Fairy   518     68      65      65      125     115     80
0.9933297291887441      Gigalith        Rock            515     85      135     130     60      80      25
grunt>
```

```
2017-10-29 12:35:11,223 [main] INFO  c
cess : 1
((Swanna,75))
((Gabite,68))
((Exeggutor,95))
((Dragonair,61))
((Shellder,30))
grunt>
```

Result of Player2.txt—

```
grunt> cat player2.txt;
0.9990708326219241      Bronzong        Steel   Psychic 500     67      89      116     79      116     33
0.9990272659534193      Crawdaunt       Water   Dark    468     63      120     85      90      55      55
0.9987790959169232      Exploud Normal          490     104     91      63      91      73      68
0.9928881560154161      Bronzor Steel   Psychic 300     57      24      86      24      86      23
0.9927841521844853      Reuniclus       Psychic         490     110     65      75      125     85      30
grunt>
```

```
2017-10-29 12:37:54,966 [main] INFO  o
cess : 1
((Gothitelle,70))
((Slowking,95))
((Scizor,70))
((Gliscor,75))
((Nidoqueen,90))
grunt>
```