Problem 1:

1. Dataset:
   P.csv : Generate Point (1200000 points)
   R.csv : Generate Rectangle(5000000 rectangles)

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
P.csv	file	347.58 MB	3	64 MB	2018-09-30 21:01	rw-r--r--	hadoop	hadoop
R.csv	file	135.14 MB	3	64 MB	2018-10-03 19:50	rw-r--r--	hadoop	hadoop

File display in Hadoop

```
1,6649,8301
2,3722,588
3,5782,886
4,9381,8613
5,8744,1534
6,5734,7961
7,3693,1470
8,5498,7262
9,8134,9649
10,1696,4954
11,7808,8367
12,9284,715
13,4542,6228
14,4269,3936
15,4565,176
16,4639,6523
17,5155,1191
18,3229,8398
19,765,9272
20,2783,4786
21,2884,7371
22,2923,6487
23,7891,1860
24,9843,7110
25,6392,1796
26,3083,7122
```

Example data in P.csv

```
r1,9083,8413,6,4
r2,6632,746,11,1
r3,2364,5604,10,3
r4,6195,1678,18,1
r5,9803,2658,15,4
r6,9876,7363,19,5
r7,788,4641,3,4
r8,8339,1083,17,2
r9,5114,6063,13,1
r10,7731,5933,2,5
r11,7084,4010,19,5
r12,2302,6208,7,5
r13,8581,162,14,4
r14,7467,4940,18,4
r15,9861,3401,7,4
r16,8921,9583,11,5
r17,9034,9590,18,3
r18,2451,8486,1,5
r19,6396,8157,10,2
r20,3323,5775,20,2
r21,3978,1896,1,4
r22,6192,4205,7,1
r23,3679,7329,3,5
r24,5210,3123,13,5
r25,9377,9444,16,2
r26,58,4311,1,2
```

Example data in R.csv

2. MapReduce Job:
   1 job, which contain 2 mappers(one for point filter and one for rectangle filter) and 1 reducer.
   Main logic: divide whole (0,0,10000,10000) area into 20 * 20 sub area and each area's index as key for rectangles and points.
   Execute command example:
   hadoop jar ./prob1.jar prob1 /user/hadoop/HW-2/R.csv /user/hadoop/HW-2/P.csv /user/hadoop/HW-2/output1 100 500 1000 2000
   Example result:(example from result)

```
r3529194        <r3529194,(253,901)>
r3529194        <r3529194,(261,903)>
r3529194        <r3529194,(264,902)>
r3529194        <r3529194,(263,904)>
r3529194        <r3529194,(265,901)>
r3529194        <r3529194,(265,902)>
r3529194        <r3529194,(258,901)>
r3529194        <r3529194,(264,902)>
r3529194        <r3529194,(265,901)>
r3529194        <r3529194,(267,902)>
r3529194        <r3529194,(256,901)>
r3529194        <r3529194,(261,902)>
r3529194        <r3529194,(264,904)>
r4815579        <r4815579,(202,741)>
r4815579        <r4815579,(197,741)>
r4815579        <r4815579,(200,741)>
r3843847        <r3843847,(212,788)>
r3843847        <r3843847,(207,788)>
r3843847        <r3843847,(202,788)>
r3843847        <r3843847,(215,788)>
r3843847        <r3843847,(218,787)>
r3843847        <r3843847,(206,788)>
r3843847        <r3843847,(211,788)>
r3843187        <r3843187,(474,694)>
r3843187        <r3843187,(483,694)>
r3843187        <r3843187,(487,695)>
```

Problem2

1. Dataset:
   Provided.

airfield.json	file	944.94 KB	3	64 MB	2018-10-04 18:14	rw-r--r--	hadoop	hadoop

2. Result:
   Execute command example:
   hadoop jar ./p2.jar pro2.p2 /user/hadoop/HW-2/temp/airfield.json /user/hadoop/HW-2/ourput1

```
1024    1320,1
1025    1312,32
1026    940,9
1027    1752,84
1028    2499,1
1029    2399,3
1032    1710,21
2052    99,99
256     1889,1889
258     2314,37
259     548,155
260     1139,1
32      2006,1
```

```
256     1889,1889
258     2314,37
259     548,155
260     1139,1
32      2006,1
33      70,70
34      366,41
36      252,4
40      1137,15
64      1859,3
66      110,110
72      67,31
```

Problem3:

1. Dataset:(1200000 points)

P.csv	file	111.91 MB	3	64 MB	2018-10-04 19:08	rw-r--r--	hadoop	hadoo

Example Data:(delete the index of every point and keep everything else the same)

```
7745,264
2580,3713
8267,681
4702,1088
1658,2851
4513,8374
1771,6199
8182,8796
2056,8958
473,7045
3351,6257
876,5187
7588,2307
6918,2855
1753,6143
1732,2755
9161,4857
3535,846
3164,6276
254,9523
6116,416
6386,3654
1071,3629
7391,325
5781,9488
1339,3014
```

2. Result:(k = 20 , R = 5)

   Error execute command example:

```
hadoop@hadoop-VirtualBox:~/Documents/HW-2_3$ hadoop jar ./prob3.jar prob3  /user
/hadoop/HW-2/P.csv /user/hadoop/HW-2/test4 5
Warning: $HADOOP_HOME is deprecated.

Usage:wordcount <HDFS input files><HDFS output file><R><K>
```

   Standard execute command example:

   hadoop jar ./prob3.jar prob3 /user/hadoop/HW-2/P.csv /user/hadoop/HW-2/test3 5 20

   Example Data:

```
35,1,1   k=14
42,19,1  k=16
5,44,1   k=17
17,15,1  k=19
41,6,1   k=17
47,47,1  k=14
49,23,1  k=19
25,4,1   k=12
4,36,1   k=16
1,44,1   k=17
48,33,1  k=19
21,6,1   k=19
4,42,1   k=17
20,23,1  k=16
32,45,1  k=14
35,21,1  k=15
39,11,1  k=17
12,6,1   k=18
31,7,1   k=17
45,17,1  k=16
39,4,1   k=12
34,49,1  k=16
17,49,1  k=18
3,6,1    k=18
10,21,1  k=18
2,14,1   k=6
```

```
4,87,1   k=17
37,86,1  k=16
10,89,1  k=19
29,73,1  k=19
43,52,1  k=13
3,50,1   k=12
7,53,1   k=17
33,81,1  k=18
49,63,1  k=17
14,65,1  k=18
41,81,1  k=18
18,61,1  k=19
45,97,1  k=17
46,58,1  k=18
3,64,1   k=19
42,92,1  k=19
4,91,1   k=18
30,93,1  k=19
36,84,1  k=17
23,97,1  k=18
37,68,1  k=19
2,73,1   k=9
8,53,1   k=18
23,96,1  k=19
47,80,1  k=18
37,96,1  k=19
```