1. Show number of downloads for package ggplot2.

```java
import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class GGPlot2DownloadCounter {

  public static class GGPlot2Mapper extends Mapper<LongWritable, Text, Text, IntWritable> {
    private final static IntWritable one = new IntWritable(1);
    private Text packageName = new Text("ggplot2");

    @Override
    protected void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException {
      // Assuming CSV structure: date,time,size,r_version,r_arch,r_os,package,version,country,ip_id
      String[] parts = value.toString().split(",");
      // Check if parts length is correct and package name matches "ggplot2"
      if (parts.length > 7 && parts[6].trim().equalsIgnoreCase("\"ggplot2\"")) {
        context.write(packageName, one); // Emit package name with count of 1
      }
    }
  }

  public static class SumReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
    @Override
    protected void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException, InterruptedException {
      int sum = 0;
      for (IntWritable val : values) {
        sum += val.get();
      }
      context.write(key, new IntWritable(sum)); // Emit total count for package
    }
  }

  public static void main(String[] args) throws Exception {
    if (args.length != 2) {
      System.err.println("Usage: GGPlot2DownloadCounter <input path> <output path>");
      System.exit(-1);
    }

    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, "GGPlot2 Download Counter");
    job.setJarByClass(GGPlot2DownloadCounter.class);
    job.setMapperClass(GGPlot2Mapper.class);
    job.setCombinerClass(SumReducer.class);
    job.setReducerClass(SumReducer.class);

    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(IntWritable.class);

    FileInputFormat.addInputPath(job, new Path(args[0]));
    FileOutputFormat.setOutputPath(job, new Path(args[1]));

    System.exit(job.waitForCompletion(true) ? 0 : 1);
  }
}
```

```
Number of ggplot2 downloads: 83820
```

2. List the highest number of downloads by a country.

```
1   import java.io.IOException;
2   import org.apache.hadoop.conf.Configuration;
3   import org.apache.hadoop.fs.Path;
4   import org.apache.hadoop.io.IntWritable;
5   import org.apache.hadoop.io.Text;
6   import org.apache.hadoop.mapreduce.Job;
7   import org.apache.hadoop.mapreduce.Mapper;
8   import org.apache.hadoop.mapreduce.Reducer;
9   import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
10  import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
11
12  public class Query2 {
13
14      public static class CountryMapper extends Mapper<Object, Text, Text, IntWritable> {
15          private final static IntWritable one = new IntWritable(1);
16          private Text country = new Text();
17
18          @Override
19          protected void map(Object key, Text value, Context context) throws IOException, InterruptedException {
20              // Assuming CSV structure: date,time,size,r_version,r_arch,r_os,package,version,country,ip_id
21              String[] parts = value.toString().split(",");
22              if (parts.length > 8 && parts[6].replaceAll("\"", "").equals("ggplot2")) {
23                  country.set(parts[8].replaceAll("\"", ""));
24                  context.write(country, one);
25              }
26          }
27      }
28
29      public static class SumReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
30          private IntWritable result = new IntWritable();
31
32          @Override
33          protected void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException, InterruptedException {
34              int sum = 0;
35              for (IntWritable val : values) {
36                  sum += val.get();
37              }
38              result.set(sum);
39              context.write(key, result);
40          }
41      }
42
43      public static void main(String[] args) throws Exception {
44          if (args.length != 2) {
45              System.err.println("Usage: GGPlot2CountryDownloadCounter <input path> <output path>");
46              System.exit(-1);
47          }
48
49          Configuration conf = new Configuration();
50          Job job = Job.getInstance(conf, "GGPlot2 Country Download Counter");
51          job.setJarByClass(Query2.class);
52          job.setMapperClass(CountryMapper.class);
53          job.setCombinerClass(SumReducer.class);
54          job.setReducerClass(SumReducer.class);
55          job.setOutputKeyClass(Text.class);
56          job.setOutputValueClass(IntWritable.class);
57          FileInputFormat.addInputPath(job, new Path(args[0]));
58          FileOutputFormat.setOutputPath(job, new Path(args[1]));
59
60          System.exit(job.waitForCompletion(true) ? 0 : 1);
61      }
62  }
63
64
```

```
Country with the highest number of ggplot2 downloads: GB (36351 downloads)
```

3. What were the top 10 most popular packages?

```java
1    import java.io.IOException;
2    import org.apache.hadoop.conf.Configuration;
3    import org.apache.hadoop.fs.Path;
4    import org.apache.hadoop.io.IntWritable;
5    import org.apache.hadoop.io.LongWritable;
6    import org.apache.hadoop.io.Text;
7    import org.apache.hadoop.mapreduce.Job;
8    import org.apache.hadoop.mapreduce.Mapper;
9    import org.apache.hadoop.mapreduce.Reducer;
10   import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
11   import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
12
13   public class Query3 {
14
15       public static class PackageMapper extends Mapper<LongWritable, Text, Text, IntWritable> {
16           private final static IntWritable one = new IntWritable(1);
17           private Text packageName = new Text();
18
19           @Override
20           protected void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException {
21               // Assuming CSV structure without handling commas within fields
22               String[] parts = value.toString().split(",");
23               if (parts.length > 6 && !parts[6].trim().equals("package")) { // Skip header
24                   String packageRaw = parts[6].trim().replaceAll("^\"|\"$", ""); // Remove potential quotes
25                   packageName.set(packageRaw);
26                   context.write(packageName, one);
27               }
28           }
29       }
30
31       public static class SumReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
32           private IntWritable result = new IntWritable();
33
34           @Override
35           public void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException, InterruptedException {
36               int sum = 0;
37               for (IntWritable val : values) {
```

```java
37               for (IntWritable val : values) {
38                   sum += val.get();
39               }
40               result.set(sum);
41               context.write(key, result);
42           }
43       }
44
45       public static void main(String[] args) throws Exception {
46           if (args.length != 2) {
47               System.err.println("Usage: PackagePopularityCounter <input path> <output path>");
48               System.exit(-1);
49           }
50
51           Configuration conf = new Configuration();
52           Job job = Job.getInstance(conf, "Package Popularity Counter");
53           job.setJarByClass(Query3.class);
54           job.setMapperClass(PackageMapper.class);
55           job.setCombinerClass(SumReducer.class);
56           job.setReducerClass(SumReducer.class);
57           job.setOutputKeyClass(Text.class);
58           job.setOutputValueClass(IntWritable.class);
59
60           FileInputFormat.addInputPath(job, new Path(args[0]));
61           FileOutputFormat.setOutputPath(job, new Path(args[1]));
62
63           if (!job.waitForCompletion(true)) {
64               System.exit(1);
65           }
66       }
67   }
68
```

```
Top 10 most popular packages:
package
ragg            85238
ggplot2         83820
textshaping     83090
rlang           64202
cli             56377
vctrs           53103
dplyr           52822
lifecycle       51680
devtools        45150
glue            43576
```

4. What is the most popular package in Ireland?

```java
1   import java.io.IOException;
2   import org.apache.hadoop.conf.Configuration;
3   import org.apache.hadoop.fs.Path;
4   import org.apache.hadoop.io.IntWritable;
5   import org.apache.hadoop.io.LongWritable;
6   import org.apache.hadoop.io.Text;
7   import org.apache.hadoop.mapreduce.Job;
8   import org.apache.hadoop.mapreduce.Mapper;
9   import org.apache.hadoop.mapreduce.Reducer;
10  import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
11  import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
12
13  public class Query4 {
14
15      public static class PackageMapper extends Mapper<LongWritable, Text, Text, IntWritable> {
16          private final static IntWritable one = new IntWritable(1);
17          private Text packageText = new Text();
18
19          @Override
20          protected void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException {
21              // Convert the record (CSV line) into a string array
22              String[] line = value.toString().split(",");
23
24              // Ensure line has enough parts to avoid IndexOutOfBoundsException
25              if (line.length > 9) {
26                  // Filter for records where country is "IE"
27                  if ("\"IE\"".equals(line[8].trim())) {
28                      // Remove quotes from package name and emit
29                      String packageName = line[6].replaceAll("\"", "");
30                      packageText.set(packageName);
31                      context.write(packageText, one);
32                  }
33              }
34          }
35      }
36
37      public static class SumReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
38          @Override
39          protected void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException, InterruptedException {
40              int sum = 0;
41              for (IntWritable val : values) {
42                  sum += val.get();
43              }
44              context.write(key, new IntWritable(sum));
45          }
46      }
47
48      public static void main(String[] args) throws Exception {
49          if (args.length != 2) {
50              System.err.println("Usage: MostPopularPackageIreland <input path> <output path>");
51              System.exit(-1);
52          }
53
54          Configuration conf = new Configuration();
55          Job job = Job.getInstance(conf, "Most Popular Package in Ireland");
56          job.setJarByClass(Query4.class);
57          job.setMapperClass(PackageMapper.class);
58          job.setReducerClass(SumReducer.class);
59          job.setOutputKeyClass(Text.class);
60          job.setOutputValueClass(IntWritable.class);
61
62          FileInputFormat.addInputPath(job, new Path(args[0]));
63          FileOutputFormat.setOutputPath(job, new Path(args[1]));
64
65          System.exit(job.waitForCompletion(true) ? 0 : 1);
66      }
67  }
68
```

```
The most popular package in Ireland is: ggplot2 with 186 downloads
```

5. What OS is most popular among the R programmers?

```java
import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class Query5 {

    public static class OSCountMapper extends Mapper<LongWritable, Text, Text, IntWritable> {
        private final static IntWritable one = new IntWritable(1);
        private Text os = new Text();

        @Override
        protected void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException {
            // Split CSV line. Note: This doesn't handle commas within fields.
            String[] parts = value.toString().split(",");
            if (parts.length > 5 && !parts[3].equals("NA")) {
                // Extract OS, assuming quotes are used.
                String osName = parts[5].replaceAll("^\"|\"$", "");
                os.set(osName);
                context.write(os, one);
            }
        }
    }

    public static class SumReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
        @Override
        public void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException, InterruptedException {
            int sum = 0;
            for (IntWritable val : values) {
                sum += val.get();
            }
            context.write(key, new IntWritable(sum));
        }
    }

    public static void main(String[] args) throws Exception {
        if (args.length != 2) {
            System.err.println("Usage: MostPopularOSAmongRProgrammers <input path> <output path>");
            System.exit(-1);
        }

        Configuration conf = new Configuration();
        Job job = Job.getInstance(conf, "Most Popular OS Among R Programmers");
        job.setJarByClass(Query5.class);
        job.setMapperClass(OSCountMapper.class);
        job.setCombinerClass(SumReducer.class);
        job.setReducerClass(SumReducer.class);
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);

        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));

        System.exit(job.waitForCompletion(true) ? 0 : 1);
    }
}
```

```
The most popular OS among R programmers is: mingw32 with 2489506 users
```