Exercise 1.3: Hadoop Java API

In this exercise you will write a MapReduce job in Java which computes the
average size of M&Ms per bag.
You can duplicate the color counting code and modify it.
To execute your job, run the following commands in the VM (after having
deployed your jar to the VM).

```
hadoop jar hello-hadoop.jar de.inovex.academy.hadoop.candy.AvgSizeDriver \ 
<indir> <outdir>
```

1. Form groups of 2-3 people.

2. Define the task of the mapper and the reducer. How is the work divided among them? What are the types of their input and output keys and values?

3. Implement the mapper and reducer classes.

4. Implement the driver class.

5. What might be an appropriate number of reduce tasks for this job? (Hint: there is not clear-cut answer to this.)

6. What might be an appropriate number of map tasks for this job? (Hint: this is a trick question.)

7. Deploy your program to the VM and run it (see command line above)

8. Check the correctness of the results for a tiny subset of the data (e.g., one bag).

Hints for debugging:

1. Disabling the reduce phase (by setting the number of reducers to zero) allows you to inspect the mapper output.

2. Replacing the IntSumReducer by the identity reducer allows you to inspect the (sorted) reducer input.