Name: _____ USC ID: _____

INF 551 – Fall 2017 (Afternoon section)

Quiz 12: Hadoop MapReduce (10 points), 15 minutes

Consider a sales data stored as a csv file (fields are category, state, year, and amount):

Cell,CA,2017,100
Cell,NY,2017,150
Cell,CA,2018,600
Cell,NY,2018,500
Cell,CA,2019,300

Consider computing the answer to the following SQL query using Hadoop MapReduce.

```
Select category, state, sum(amount)
From sales
Where year = 2017 or year = 2018
Group by category, state
Having min(amount) > 50
```

1. [5 points] Write the logic of **map** function in pseudocode. What will the map function output for the data above?

```
// key: line offset; value: line content split by comma
For each field in value:
 if year is 2017 or 2018:
output ((category, state), amount)
Output:
    (Cell, CA), 100
    (Cell, NY), 150
    (Cell, CA), 600
    (Cell, NY), 500
```

2. [5 points] Write the logic of **reduce** function in pseudocode. What will the reduce function output for the data above?

```
// key: (category, state); values: an iterator over amount
 sum = 0
minV= a large number
for each amount in values:
    sum += amount
    if amount<minV:
            minV=amount
if minV>50:
    output (key, sum)
Output:
    (Cell, CA), 700
    (Cell, NY), 650
```