The Means vs Variances plot was generated through the following steps:

1. The *groups.txt* file was downloaded using the EMR Amazon web interface onto my local linux machine.

2. The file was '*scp*'ed to hadoop@ec2-54-218-219-209.us-west-2.compute.amazonaws.com:~/ganesh/

3. A table with fields of datatype string was created using the hive command:

```
hive> create table group_value_string (group string, value string) row format delimited fields terminated by '\t' lines terminated by '\n' stored as textfile;
```

4. Data was loaded into the table from *groups.txt*, now present in the folder ~/ganesh in the local file system:

```
hive> load data local inpath 'groups.txt' into table group_value_string;
```

5. Table for the actual data analysis:

```
hive>create table group_value (group int, value double) row format delimited fields terminated by '\t' lines terminated by '\n';
```

6. 'groups' table being populated using:

```
hive>insert into table group_value select cast(group as int), cast(value as double) from group_value_string;
```

7. creating tables for the group means and group variances:

```
hive>create table group_means (group int, mean double) row format delimited fields terminated by '\t' lines terminated by '\n';
hive>create table group_variances (group int, variance double) row format delimited fields terminated by '\t' lines terminated by '\n';
```

8. using 'group by' to compute the group means and variances :

```
hive>insert into table group_means select gv.group, avg(gv.value) from group_value gv group by gv.group;
hive>insert into table group_variances select gv.group, variance(gv.value) from group_value gv group by gv.group;
```

9. Writing the group mean and group variance tables into a text file on the local system with '\t' field delimiter:

```
hive>insert overwrite local directory 'means' select concat_ws('\t', cast(group as string), cast(mean as string)) from group_means;
hive>insert overwrite local directory 'variances' select concat_ws('\t', cast(group as string), cast(variance as string)) from group_variances;
```

10. The text files with group means and variances were '*scp*'ed back to my local linux machine (for the final plot) :

scp -i GaneshKeyPair.pem hadoop@ec2-54-218-219-209.us-west-2.compute.amazonaws.com:~/ganesh/means/* means.txt

scp -i GaneshKeyPair.pem hadoop@ec2-54-218-219-209.us-west-2.compute.amazonaws.com:~/ganesh/variances/* vars.txt