Assignment 17.2

1. **Read the text file, and create a tupled rdd.**

```
scala> val read = sc.textFile("/home/acadgild/dataset.txt")
read: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[23] at textFile at <co
nsole>:27

scala> read.count()
res15: Long = 22

scala>
```

2. **Find the count of total number of rows present.**

```
scala> val read = sc.textFile("/home/acadgild/dataset.txt")
read: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[23] at textFile at <co
nsole>:27

scala> read.count()
res15: Long = 22

scala>
```

3. **What is the distinct number of subjects present in the entire school**

```
scala> case class College(name:String,subject:String,grades:String,marks:Int)
defined class College

scala> val college = read map{l => {
val a = l.split(",")
College(a(0),a(1),a(2),a(3).toInt)
}
}
college: org.apache.spark.rdd.RDD[College] = MapPartitionsRDD[32] at map at <con
sole>:31

scala> college.collect
res25: Array[College] = Array(College(Mathew,science,grade-3,45), College(Mathew
,history,grade-2,55), College(Mark,maths,grade-2,23), College(Mark,science,grade
-1,76), College(John,history,grade-1,14), College(John,maths,grade-2,74), Colleg
e(Lisa,science,grade-1,24), College(Lisa,history,grade-3,86), College(Andrew,mat
hs,grade-1,34), College(Andrew,science,grade-3,26), College(Andrew,history,grade
-1,74), College(Mathew,science,grade-2,55), College(Mathew,history,grade-2,87),
College(Mark,maths,grade-1,92), College(Mark,science,grade-2,12), College(John,h
istory,grade-1,67), College(John,maths,grade-1,35), College(Lisa,science,grade-2
,24), College(Lisa,history,grade-2,98), College(Andrew,maths,grade-1,23), Colleg
e(Andrew,science,grade-3,44), College(Andrew,history,grade-2,77))

scala>
```

```
scala> val subject = college.map(c => c.subject)
subject: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[33] at map at <cons
ole>:33

scala> subject.collect
res26: Array[String] = Array(science, history, maths, science, history, maths, s
cience, history, maths, science, history, science, history, maths, science, hist
ory, maths, science, history, maths, science, history)

scala>
```

```
scala> val uniquesubject = subject.distinct
uniquesubject: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[36] at distin
ct at <console>:35

scala> uniquesubject.collect
res27: Array[String] = Array(maths, history, science)

scala> uniquesubject.count()
res28: Long = 3

scala>
```

4. **What is the count of the number of students in the school, whose name is Mathew and marks is 55**

```
scala> val filter = college filter(c => (c.name == "Mathew" && c.marks == 55))
filter: org.apache.spark.rdd.RDD[College] = MapPartitionsRDD[38] at filter at <c
onsole>:33

scala> filter.collect
res30: Array[College] = Array(College(Mathew,history,grade-2,55), College(Mathew
,science,grade-2,55))

scala> filter.count()
res31: Long = 2
```