1. The following variables are included in the Assignment Data 1 dataset:

- ID
- Last Name
- First Name
- City
- State
- Gender
- Student Status
- Major
- Country
- Age
- SAT
- Average score (grade)
- Height (in)
- Newspaper readership (times/wk)

2. The dataset has biographical information about students across different countries. The dataset seems to be mapping the relationship between SAT scores or reading hours per week (possible dependent variables) and a number of other variables.

Variable Name	**Format**	**Type**	**Description**
ID	Integer	Discrete	Identification number
Last Name	String	Categorical	Surname of student
First Name	String	Categorical	First name of student
City	String	Categorical	Home city of student
State	String	Categorical	Home state of student
Gender	String	Categorical	Sex of student
Student Status	String	Categorical	Whether student is graduate or undergraduate
Major	String	Categorical	Major of student
Country	String	Categorical	Home country of student
Age	Integer	Discrete	Age of student
SAT	Integer	Discrete	SAT score of student

Average score (grade)	Integer	Discrete	Score of student in an unspecified course or set of courses
Height (in)	Numeric	Continuous	Height of student in inches
Newspaper readership (times/wk)	Integer	Discrete	Amount of hours student reads the newspaper each week

3. This is sample data (subset of the population) since we can assume that this is not the entire population of students that attend this university.
4. There are 15 men and 15 women in the sample.

```
> #Calculating number of men and women
> sum(data_assignment1$Gender == 'Male')
[1] 15
> sum(data_assignment1$Gender == 'Female')
[1] 15
```

5. The average age is 25.2 years in the sample.

```
> #Average Age
> mean(data_assignment1$Age)
[1] 25.2
>
```

6. There are 15 graduate and 15 undergraduate students in the sample.

```
> sum(data_assignment1$`Student Status` == 'Graduate')
[1] 15
> sum(data_assignment1$`Student Status` == 'Undergraduate')
[1] 15
```

7. The average SAT score is 1849 across all students. However, the average score is not the same for graduates (1841) and undergraduates (1857).

```
> mean(data_assignment1$SAT)
[1] 1848.9
```

```
> aggregate(x = data_assignment1$SAT,                # Specify data column
+           by = list(data_assignment1$`Student Status`),              # Specify group indicator
+           FUN = mean)                            # Specify function (i.e. mean)
        Group.1      x
1      Graduate 1841.2
2 Undergraduate 1856.6
```

8. Men read the paper less than women. Men read the paper at a rate of 4.5x/week while women read the paper at a rate of 5.2x/week.

```
> aggregate(x = data_assignment1$`Newspaper readership (times/wk)`,                # Specify data c
olumn
+           by = list(data_assignment1$Gender),              # Specify group indicator
+           FUN = mean)                            # Specify function (i.e. mean)
  Group.1        x
1  Female 5.200000
2    Male 4.533333
```

9. See the measure of location and dispersion for all quantitative variables below:

```
> summary(data_assignment1)
       ID         Last Name          First Name
 Min.   : 1.00   Length:30          Length:30
 1st Qu.: 8.25   Class :character   Class :character
 Median :15.50   Mode  :character   Mode  :character
 Mean   :15.50
 3rd Qu.:22.75
 Max.   :30.00
    City              State              Gender
 Length:30          Length:30          Length:30
 Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character



 Student Status        Major              Country
 Length:30          Length:30          Length:30
 Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character



     Age            SAT       Average score (grade)  Height (in)
 Min.   :18.0   Min.   :1338   Min.   :63.00         Min.   :59.00
 1st Qu.:19.0   1st Qu.:1658   1st Qu.:72.00         1st Qu.:63.00
 Median :23.0   Median :1817   Median :79.50         Median :66.50
 Mean   :25.2   Mean   :1849   Mean   :80.37         Mean   :66.43
 3rd Qu.:30.0   3rd Qu.:2032   3rd Qu.:88.00         3rd Qu.:70.75
 Max.   :39.0   Max.   :2309   Max.   :96.00         Max.   :75.00
 Newspaper readership (times/wk)
 Min.   :3.000
 1st Qu.:4.000
 Median :5.000
 Mean   :4.867
 3rd Qu.:6.000
 Max.   :7.000
```

Here's the R output needed to produce the script:

If:
Age <- c(22, 25, 18, 20)
Name <- c("James", "Mathew", "Olivia", "Stella")
Gender <- c("M", "M", "F", "F")

Then: what is the R-code for getting the following output:
```
##   Age Name  Gender
## 1  22  James  M
##  2 25  Mathew M
```

Solution

```
Age <- c(22, 25, 18, 20)
Name <- c("James", "Mathew", "Olivia", "Stella")
Gender <- c("M", "M", "F", "F")

combined <- data.frame(Age, Name, Gender)
combined
combined[combined$Gender=="M", ]
```

```
> combined
  Age    Name Gender
1  22  James      M
2  25 Mathew      M
3  18 Olivia      F
4  20 Stella      F
> combined[combined$Gender=="M", ]
  Age    Name Gender
1  22  James      M
2  25 Mathew      M
```