COMP 3340 Assignment 1 Part 2)

Question 1:

Question 1A)

There is not strong correlation between variables within the students academic performance dataset, however the correlation between raisedHands and VisITedResources is the strongest with a 0.692 pearsons coefficient, which is almost defined as a strong correlation.

Table:

	raisedhands	VisITedResources	AnnouncementsView	Discussion
raisedhands	1.0000000	0.6915717	0.6439178	0.3393860
VisITedResources	0.6915717	1.0000000	0.5945000	0.2432918
AnnouncementsView	0.6439178	0.5945000	1.0000000	0.4172900
Discussion	0.3393860	0.2432918	0.4172900	1.0000000

Code:

```
question_1a <- function(datafile){
  #function takes a dataframe as input and uses the cor method which calculates the pearsons correlation for all the variable pairs.
  numeric_columns <- datafile[ names(datafile) %in% c("raisedhands","VisITedResources","AnnouncementsView","Discussion")]
  pearsons <- cor(numeric_columns, method = "pearson")
}
```

Question 1B)

(F) Students "raised hands" are more actively involved in study-related works

```
a <- cor(datafile$raisedhand, datafile$VisITedResources, method = "pearson")
b <- cor(datafile$raisedhand, datafile$AnnouncementsView, method = "pearson")
c <- cor(datafile$raisedhand, datafile$Discussion, method = "pearson")
print(a)
print(b)
print(c)
```

```
[1] 0.6915717
[1] 0.6439178
[1] 0.339386
```

(F) No apparent gender bias when it comes to subject/topic choices

```
        Pearson's Chi-squared test

data:  datafile$gender and datafile$Topic
X-squared = 23.04, df = 11, p-value = 0.01745
```

	Arabic	Biology	Chemistry	English	French	Geology	History	IT	Math	Quran	Science	Spanish
F	16	10	12	19	30	10	8	32	5	9	23	1
M	43	20	12	26	35	14	11	63	16	13	28	24

```
gender_topic <- chisq.test(datafile$gender, datafile$Topic)
print(gender_topic)
```

(T) Girls seem to have better overall performance than boys

```
        Pearson's Chi-squared test

data:  datafile$gender and datafile$Class
X-squared = 33.326, df = 2, p-value = 5.798e-08
```

(F) Boys are generally a bit more open to discussions, visiting resources, and raising hands

Means in Contribution by Gender.

```
gender_contribution <- split(X <- datafile[ names(datafile) %in% c("raisedhands","Discussion","VisITedResources","gender")], X$gender)
i = 1
while(i < 3){
  X <- colMeans(gender_contribution[[i]][sapply(gender_contribution[[i]], is.numeric)], na.rm=TRUE)
  print(gender_contribution[[i]]$gender[1])
  print(X)
  i = i +1
}
```

```
[1] "F"
     raisedhands VisITedResources       Discussion
        52.86286         64.00000         47.82857
[1] "M"
     raisedhands VisITedResources       Discussion
        43.28197         49.51803         40.67541
```

(T) Those who participated more (higher counts in Discussion, Announcement Views,

Raised Hands), usually perform better.

Means in contribution values by Results.

```
[1] "H"                 "70.2887323943662" "53.3802816901408" "53.6619718309859"
[1] "L"                 "16.8897637795276" "15.5748031496063" "30.8346456692913"
[1] "M"                 "48.9383886255924" "40.9620853080569" "43.7914691943128"
```

```
# Determines mean values for class contribution by results
numeric_columns <- datafile[ names(datafile) %in% c("raisedhands","AnnouncementsView","Discussion","Class")]
X <- split(numeric_columns, numeric_columns$Class)
i =1;
while(i < 4)
{
  out <- c(X[[i]]$Class[1],mean(X[[i]]$raisedhands),mean(X[[i]]$AnnouncementsView),mean(X[[i]]$Discussion))
  print(out)
  i = i + 1
}
```