MAX 503 Assignment # 4

```
getwd()
```

```
## [1] "E:/MARKETING ANALYTICS/Semester 2/Marketing Research & Engineering/Assignments/Assignment 4"
```

```
ecomm.df <- read.csv("ecommerce-data.csv")
str(ecomm.df)
```

```
## 'data.frame':    1593 obs. of  45 variables:
##  $ dateTime                   : chr  "7/25/2014 14:10" "7/25/2014 15:01" "7/25/2014 15:15" "7/25/2014
##  $ country                    : chr  "United States" "United States" "United States" "United States"
##  $ city                       : chr  "Monroe" "Ambler" "Beaumont" "Cedar City" ...
##  $ region                     : chr  "LA" "PA" "TX" "UT" ...
##  $ screenRed                  : chr  "1280x1024" "1280x800" "768x1024" "1280x960" ...
##  $ surveyType                 : chr  "At Exit" "At Exit" "At Exit" "At Exit" ...
##  $ purposeProductInfo         : chr  "Products" "" "" "Products" ...
##  $ purposeBuyFromSite         : chr  "" "Buy from this site" "" "" ...
##  $ purposeComparePricing      : chr  "" "Compare pricing" "Compare pricing" "" ...
##  $ purposeInfoAndResources    : chr  "Resources" "" "" "" ...
##  $ purposeInfoOnOrder         : chr  "" "" "" "" ...
##  $ purposeOther               : chr  "" "" "" "" ...
##  $ taskFindWhatLookingFor     : chr  "" "" "" "Most or all of it" ...
##  $ concernShippingCost        : chr  "" "" "" "Shipping costs" ...
##  $ concernDeliverySpeed       : chr  "" "" "" "" ...
##  $ concernWarranties          : chr  "" "" "" "" ...
##  $ concernEaseToReturnProduct : chr  "Ease of returning (if I am not satisfied with product)" "" "" "
##  $ concernProductSafety       : chr  "" "" "" "" ...
##  $ concernRightForMyChild     : chr  "" "" "" "" ...
##  $ concernProductQuality      : chr  "Product durability/quality" "" "" "" ...
##  $ concernProductEffectiveness: chr  "Product effectiveness/will it work" "" "" "" ...
##  $ concernOther               : chr  "" "" "" "" ...
##  $ concernNone                : chr  "" "" "" "" ...
##  $ intentWasPlanningToBuy     : chr  "" "Yes" "" "" ...
##  $ profile                    : chr  "Parent" "Parent" "Parent" "Person with [condition A]" ...
##  $ whenSiteUsed               : chr  "In the past week" "In the past year" "Never. This is my first v
##  $ purchasedBefore            : chr  "Yes, once" "Yes, once" "" "" ...
##  $ purchasedWhen              : chr  "In the past month" "In the past year" "" "" ...
##  $ productKnewWhatWanted      : chr  "Yes" "Yes" "Yes" "Somewhat" ...
##  $ productSiteHasWhatWanted   : chr  "" "" "" "Yes, several of them" ...
##  $ purchaseExpectInNextMonth  : int  5 3 3 3 5 3 5 NA 5 4 ...
##  $ siteFirstHeardAbout        : chr  "In the past year" "More than 1 year ago" "Just now, from the we
##  $ age                        : chr  "25-34" "35-44" "35-44" "25-34" ...
##  $ gender                     : chr  "Female" "Female" "Female" "Female" ...
##  $ behavNumVisits             : int  13 3 2 1 1 1 4 1 2 2 ...
##  $ behavReferral              : chr  "Direct" "Unbranded Search" "Unbranded Search" "Unbranded Search
##  $ behavPageviews             : chr  "4 to 6" "1" "10+" "10+" ...
```

```
##  $ behavHomePage              : int  1 0 0 0 0 1 0 1 1 1 ...
##  $ behavDetailProdA           : int  1 0 0 1 0 1 1 0 1 1 ...
##  $ behavDetailProdB           : int  0 0 0 1 0 1 1 1 1 0 ...
##  $ behavDetailProdC           : int  0 0 0 0 0 0 1 0 1 0 ...
##  $ behavAnySolution           : int  0 0 1 1 0 0 1 0 1 0 ...
##  $ behavAnySale               : int  0 0 1 0 0 0 1 0 1 1 ...
##  $ behavCart                  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ behavConversion            : int  0 0 0 0 0 0 0 0 0 0 ...
```

1. Using the integer approximation of page views (see Exercises in Sect.4.10), describe page views for parents, teachers, and health professionals. Use a by() or aggregate() function as appropriate.

```
pageViewInt <- rep(NA, length(ecomm.df$behavPageviews))
pageViewInt[ecomm.df$behavPageviews=="0"] <- 0
pageViewInt[ecomm.df$behavPageviews=="1"] <- 1
pageViewInt[ecomm.df$behavPageviews=="2 to 3"] <- 2
pageViewInt[ecomm.df$behavPageviews=="4 to 6"] <- 4
pageViewInt[ecomm.df$behavPageviews=="7 to 9"] <- 7
pageViewInt[ecomm.df$behavPageviews=="10+"] <- 10
ecomm.df$pageViewInt <- pageViewInt

aggregate(pageViewInt ~ profile, data = ecomm.df, summary)
```

```
##                     profile pageViewInt.Min. pageViewInt.1st Qu.
## 1                         0         2.000000            2.000000
## 2      Friend/family friend         1.000000            2.000000
## 3       Health Professional         1.000000            2.000000
## 4                     Other         1.000000            2.000000
## 5                    Parent         0.000000            2.000000
## 6 Person with [condition A]         0.000000            2.000000
## 7                  Relative         1.000000            2.000000
## 8                   Teacher         0.000000            2.000000
##   pageViewInt.Median pageViewInt.Mean pageViewInt.3rd Qu. pageViewInt.Max.
## 1           2.000000         2.000000            2.000000         2.000000
## 2           7.000000         5.739130           10.000000        10.000000
## 3           4.000000         5.209386           10.000000        10.000000
## 4           4.000000         4.388060            7.000000        10.000000
## 5           7.000000         6.280928           10.000000        10.000000
## 6           4.000000         5.730769           10.000000        10.000000
## 7           4.000000         5.747664           10.000000        10.000000
## 8           4.000000         5.461883           10.000000        10.000000
```

The above output displays the summary of page views for all profiles

```
by(ecomm.df$pageViewInt, ecomm.df$profile, sum)
```

```
## ecomm.df$profile: 0
## [1] 2
## ------------------------------------------------------------
## ecomm.df$profile: Friend/family friend
## [1] 132
## ------------------------------------------------------------
```

```
## ecomm.df$profile: Health Professional
## [1] 1443
## ------------------------------------------------------------
## ecomm.df$profile: Other
## [1] 588
## ------------------------------------------------------------
## ecomm.df$profile: Parent
## [1] 4874
## ------------------------------------------------------------
## ecomm.df$profile: Person with [condition A]
## [1] 298
## ------------------------------------------------------------
## ecomm.df$profile: Relative
## [1] 615
## ------------------------------------------------------------
## ecomm.df$profile: Teacher
## [1] 1218
```

The above output displays the sum of page views for all profiles

2. What are the proportions of men and women among the various visitor profiles (teacher, parent, relative, etc.)? For this question, don't count observations where the gender is not specified as male or female.

```
ecommdf_1 <- ecomm.df[ecomm.df$gender=="Male" | ecomm.df$gender == "Female",]
with(ecommdf_1,prop.table(table(profile, gender), margin=1))
```

```
##                            gender
## profile                         Female       Male
##   Friend/family friend      1.00000000 0.00000000
##   Health Professional       0.97674419 0.02325581
##   Other                     0.95348837 0.04651163
##   Parent                    0.91428571 0.08571429
##   Person with [condition A] 0.92857143 0.07142857
##   Relative                  0.92307692 0.07692308
##   Teacher                   0.96250000 0.03750000
```

3. Considering parents, teachers, and health professionals, which group has made the most purchases recently? Answer with both descriptives and a visualization.

```
ecomm.df_2 = ecomm.df[(ecomm.df$purchasedWhen=="In the past month")
& (ecomm.df$profile=="Parent"|
ecomm.df$profile=="Health Professional"|
ecomm.df$profile=="Teacher"), ]
```

Assuming most recent data as "In the past month" since "In the past week" does not have much data

```
table(ecomm.df_2$purchasedWhen,ecomm.df_2$profile)
```

```
##
##                     Health Professional Parent Teacher
##   In the past month                  11     21       3
```

The above table describes the frequency of purchases made by Parent, Health Professional and Teacher.

As shown in this frequency table, Parent made the most purchases in the past month with 21 purchases.

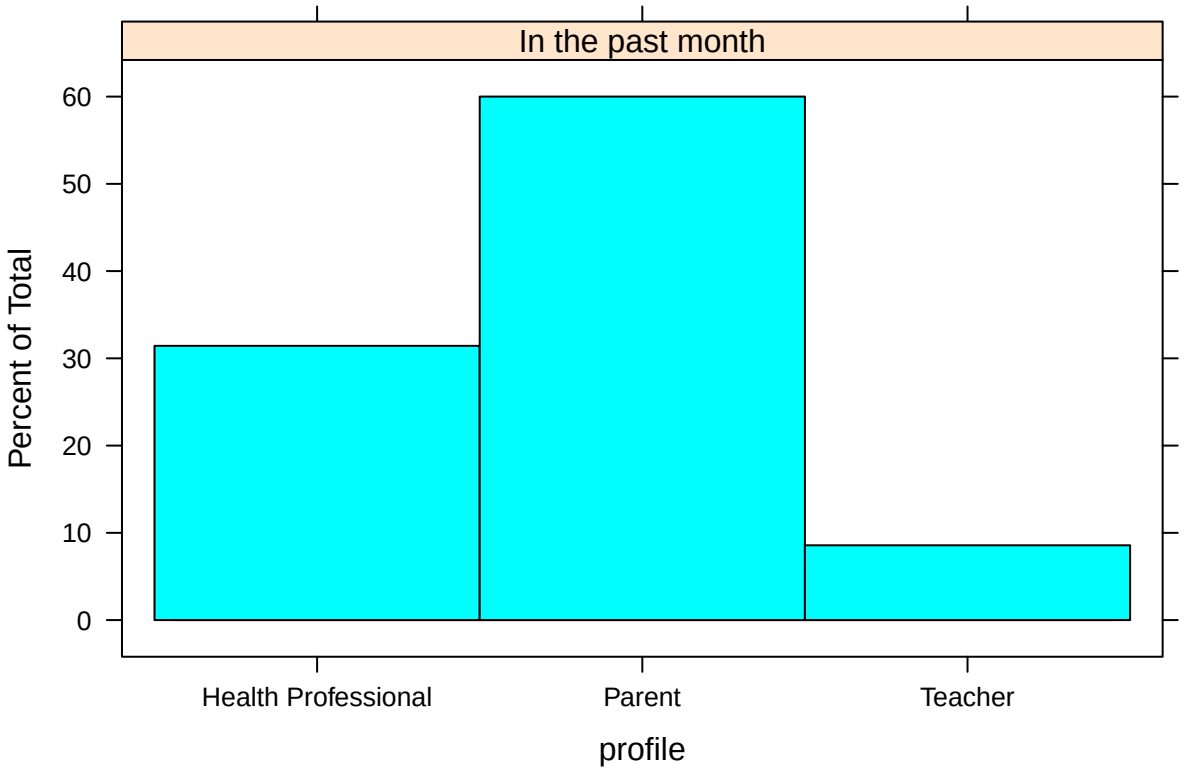
```
prop.table(table(ecomm.df_2$purchasedWhen,ecomm.df_2$profile))
```

```
##
##                     Health Professional     Parent    Teacher
##   In the past month          0.31428571 0.60000000 0.08571429
```

The above table describes the proportion of purchases made by Parent, Health Professional and Teacher.

As shown in this proportion table, Parent made 60% of purchases i.e. the most purchases in the past month.

```
library(lattice)
ecomm.df_2$purchasedWhen = as.factor(ecomm.df_2$purchasedWhen)
ecomm.df_2$profile = as.factor(ecomm.df_2$profile)
histogram(~ profile | purchasedWhen, data = ecomm.df_2)
```



4. In answering the previous question, you might use either counts or proportions. Do they give you the same answer? If not, show an example. What is a business question for which counts would be preferable?

```
table(ecomm.df_2$profile,ecomm.df_2$purchasedWhen)
```

```
##
##                         In the past month
##   Health Professional                  11
##   Parent                               21
##   Teacher                               3
```

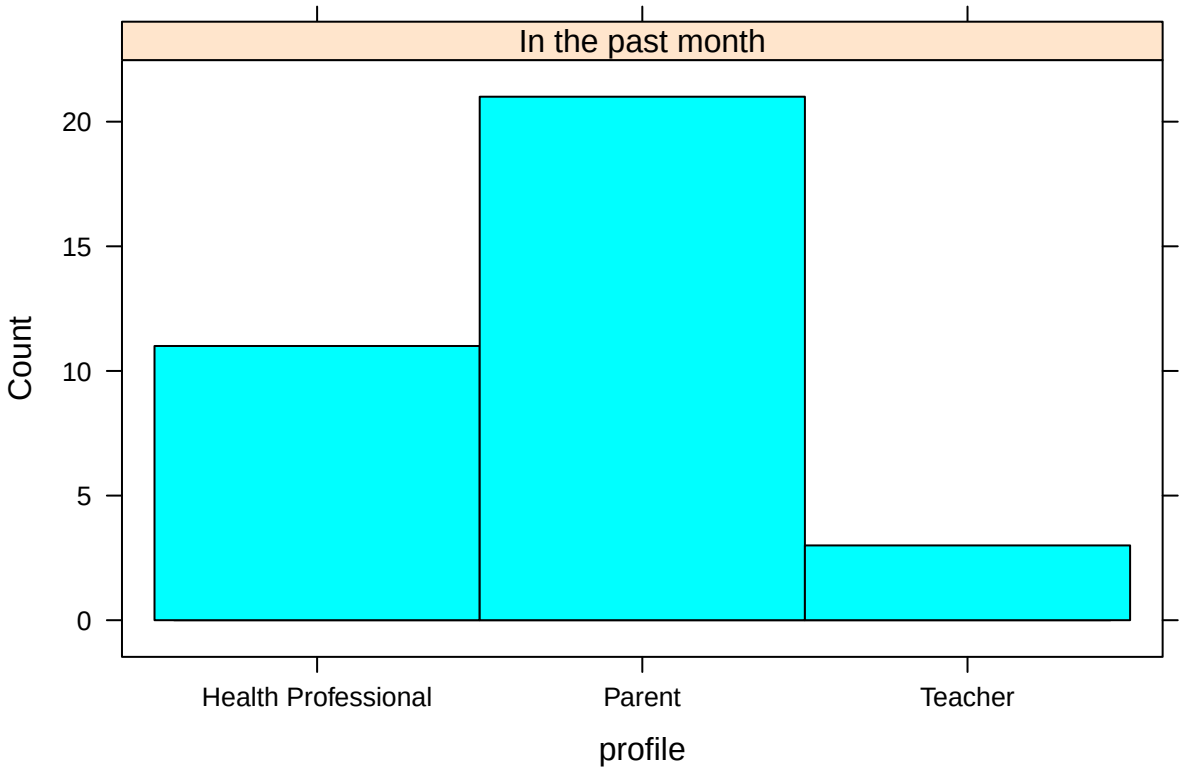
The above table gives only count, or frequency, of each value.

```
prop.table(table(ecomm.df_2$profile,ecomm.df_2$purchasedWhen),margin=2)
```

```
##
##                         In the past month
##   Health Professional           0.31428571
##   Parent                        0.60000000
##   Teacher                       0.08571429
```

The above proportion table gives proportion in relation to each column.

```
histogram(~profile | purchasedWhen, data = ecomm.df_2, type = "count")
```



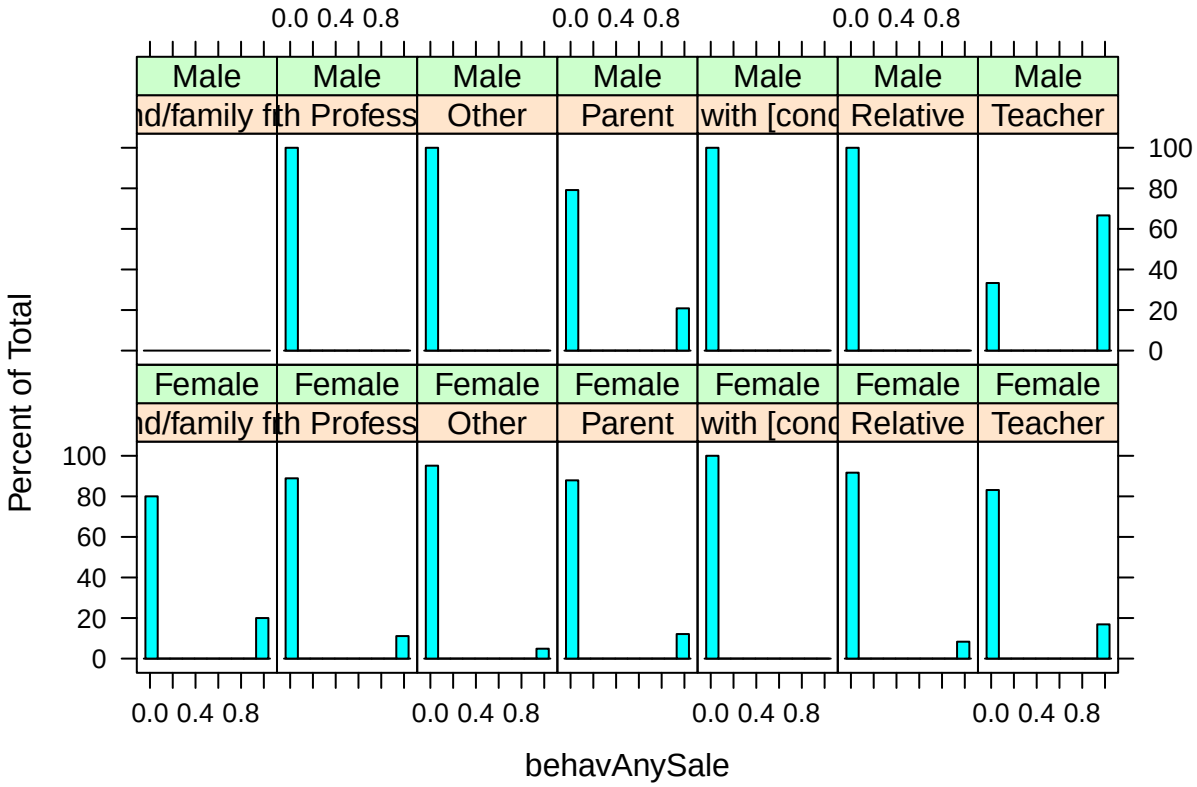
The above histograms do not give the same answer. Although they look similar, both charts represent different aspects of the output. The first histogram displayed the proportion of each profile whereas the

second histogram displayed the frequency of each profile. In other words, each histogram displays the data by different measures.

In general,for interpretation of data, proportion is preferred but to take action on business decisions, you would need the actual numbers and that is why count is preferred (for example, budgeting).

5. When we split the profiles into men and women, and consider completed purchases on the site (variable behavAnySale) which combination of profile and gender made the highest number of purchases? Which had the highest rate of purchase, relative to total number of observations?

```
library(lattice)
histogram(~behavAnySale | profile + gender, data = ecommdf_1)
```




```
aggregate(behavAnySale~profile + gender, data = ecommdf_1, sum)
```

```
##                       profile gender behavAnySale
## 1       Friend/family friend Female            1
## 2        Health Professional Female           14
## 3                      Other Female            2
## 4                     Parent Female           31
## 5  Person with [condition A] Female            0
## 6                   Relative Female            3
## 7                    Teacher Female           13
## 8        Health Professional   Male            0
## 9                      Other   Male            0
```

```
## 10                     Parent   Male            5
## 11 Person with [condition A]   Male            0
## 12                   Relative   Male            0
## 13                    Teacher   Male            2
```

Female Parent made 31 purchases which is the highest number of purchases.

```
aggregate(behavAnySale~profile + gender, data = ecommdf_1, mean)
```

```
##                       profile gender behavAnySale
## 1       Friend/family friend Female   0.20000000
## 2        Health Professional Female   0.11111111
## 3                       Other Female   0.04878049
## 4                      Parent Female   0.12109375
## 5  Person with [condition A] Female   0.00000000
## 6                    Relative Female   0.08333333
## 7                     Teacher Female   0.16883117
## 8        Health Professional   Male   0.00000000
## 9                       Other   Male   0.00000000
## 10                     Parent   Male   0.20833333
## 11 Person with [condition A]   Male   0.00000000
## 12                   Relative   Male   0.00000000
## 13                    Teacher   Male   0.66666667
```

Male Teacher had about 66.66% purchases which is the highest rate of purchases, relative to the total number of observations.