```r
getwd()
```

```
## [1] "C:/Users/Nupur Shrinet/Documents"
```

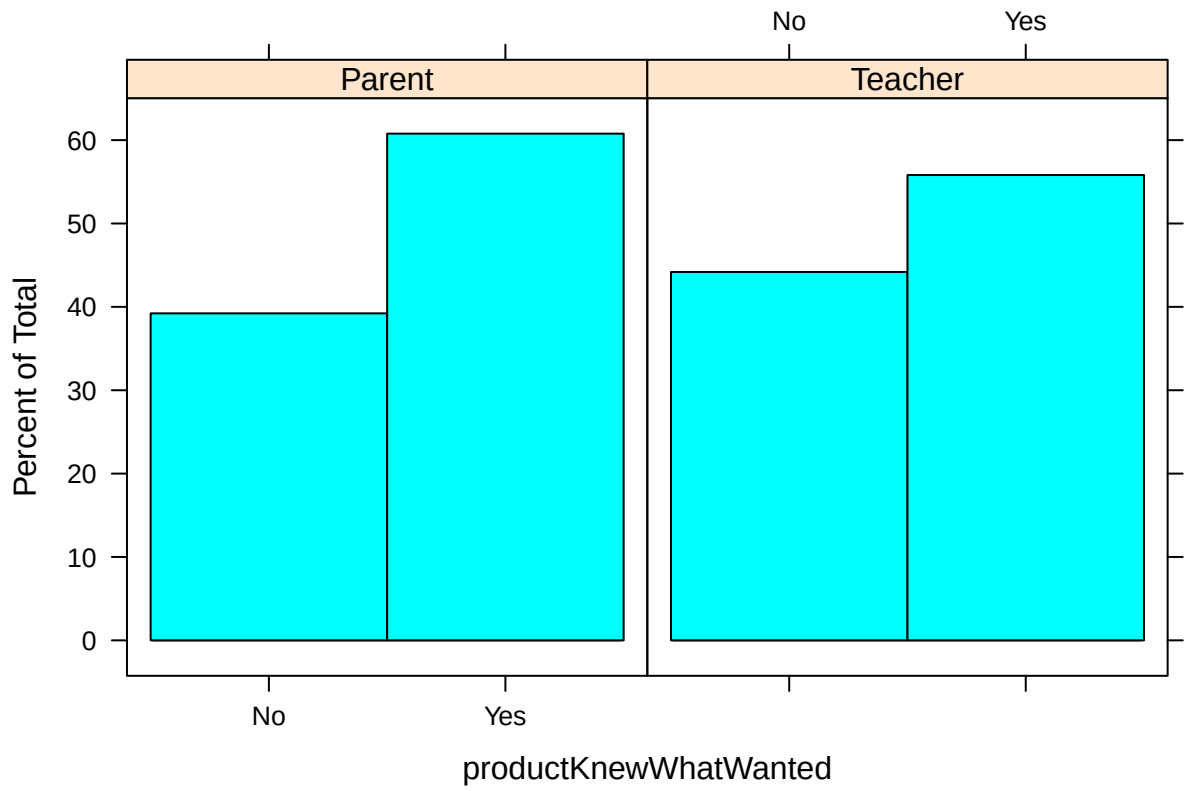
```r
ecommerce_df<-read.csv("ecommerce-data.csv")
str(ecommerce_df)
```

```
## 'data.frame':    1593 obs. of  45 variables:
##  $ dateTime                    : Factor w/ 1558 levels "7/25/2014 14:10",..: 1 2 3 4 5 6 14 7 8 9 ...
##  $ country                     : Factor w/ 44 levels "Australia","Barbados",..: 44 44 44 44 44 44 17 4
##  $ city                        : Factor w/ 980 levels "","Abilene","Abingdon",..: 563 25 76 158 132 4(
##  $ region                      : Factor w/ 110 levels "","0","1","10",..: 67 94 102 104 35 40 10 80 1(
##  $ screenRed                   : Factor w/ 91 levels "1012x569","1024x552",..: 29 34 76 35 17 43 76 7(
##  $ surveyType                  : Factor w/ 3 levels "At Arrival and Exit",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ purposeProductInfo          : Factor w/ 2 levels "","Products": 2 1 1 2 1 2 1 2 1 1 ...
##  $ purposeBuyFromSite          : Factor w/ 2 levels "","Buy from this site": 1 2 1 1 1 1 1 1 1 2 ...
##  $ purposeComparePricing       : Factor w/ 2 levels "","Compare pricing": 1 2 2 1 1 2 1 1 1 1 ...
##  $ purposeInfoAndResources     : Factor w/ 2 levels "","Resources": 2 1 1 1 2 1 2 1 1 1 ...
##  $ purposeInfoOnOrder          : Factor w/ 2 levels "","Order info": 1 1 1 1 1 1 1 1 1 1 ...
##  $ purposeOther                : Factor w/ 2 levels "","Other": 1 1 1 1 1 1 1 1 2 1 ...
##  $ taskFindWhatLookingFor      : Factor w/ 4 levels "","Most or all of it",..: 1 1 1 2 2 2 4 2 4 2 ..
##  $ concernShippingCost         : Factor w/ 2 levels "","Shipping costs": 1 1 1 2 1 1 1 1 1 1 ...
##  $ concernDeliverySpeed        : Factor w/ 2 levels "","Fast delivery": 1 1 1 1 1 1 1 1 1 1 ...
##  $ concernWarranties           : Factor w/ 2 levels "","Warranties/product guarantees": 1 1 1 1 1 1 1
##  $ concernEaseToReturnProduct  : Factor w/ 2 levels "","Ease of returning (if I am not satisfied with
##  $ concernProductSafety        : Factor w/ 2 levels "","Product safety": 1 1 1 1 1 1 1 1 1 1 ...
##  $ concernRightForMyChild      : Factor w/ 2 levels "","Whether this is right for my child": 1 1 1 1 1
##  $ concernProductQuality       : Factor w/ 2 levels "","Product durability/quality": 2 1 1 1 1 1 1 1 1
##  $ concernProductEffectiveness: Factor w/ 2 levels "","Product effectiveness/will it work": 2 1 1 1 1
##  $ concernOther                : Factor w/ 2 levels "","Other": 1 1 1 1 1 1 1 1 1 1 ...
##  $ concernNone                 : Factor w/ 2 levels "","None / no uncertainties": 1 1 1 1 1 1 1 1 2 1
##  $ intentWasPlanningToBuy      : Factor w/ 4 levels "","No","Partially (I was considering it)",..: 1 4
##  $ profile                     : Factor w/ 8 levels "0","Friend/family friend",..: 5 5 5 6 8 6 3 4 8 5
##  $ whenSiteUsed                : Factor w/ 6 levels "","In the past month",..: 3 4 6 6 6 6 3 3 6 6 ..
##  $ purchasedBefore             : Factor w/ 4 levels "","No","Yes, more than once",..: 4 4 1 1 1 1 2 4
##  $ purchasedWhen               : Factor w/ 5 levels "","In the past month",..: 2 4 1 1 1 1 1 2 1 1 ..
##  $ productKnewWhatWanted       : Factor w/ 4 levels "","No","Somewhat",..: 4 4 4 3 1 3 1 1 1 4 ...
##  $ productSiteHasWhatWanted    : Factor w/ 5 levels "","No","Not sure",..: 1 1 1 5 1 5 1 1 1 5 ...
##  $ purchaseExpectInNextMonth   : int  5 3 3 3 5 3 5 NA 5 4 ...
##  $ siteFirstHeardAbout         : Factor w/ 6 levels "","In the past hour",..: 4 6 5 2 5 2 3 1 5 1 ...
##  $ age                         : Factor w/ 9 levels "","18-24","25-34",..: 3 4 4 3 6 2 6 1 5 1 ...
##  $ gender                      : Factor w/ 4 levels "","Female","Male",..: 2 2 2 2 2 4 2 1 2 1 ...
##  $ behavNumVisits              : int  13 3 2 1 1 1 4 1 2 2 ...
##  $ behavReferral               : Factor w/ 9 levels "","Branded Search",..: 3 9 9 9 6 8 3 9 6 9 ...
##  $ behavPageviews              : Factor w/ 6 levels "0","1","10+",..: 5 2 3 3 2 3 3 5 3 6 ...
##  $ 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 ...
```

```
with(subset(ecommerce_df, profile %in% c("Parent", "Teacher") &
productKnewWhatWanted %in% c("No", "Yes")), table(profile,
productKnewWhatWanted))
```

```
##                              productKnewWhatWanted
## profile                            No Somewhat Yes
##   0                             0   0        0   0
##   Friend/family friend          0   0        0   0
##   Health Professional           0   0        0   0
##   Other                         0   0        0   0
##   Parent                        0 111        0 172
##   Person with [condition A]     0   0        0   0
##   Relative                      0   0        0   0
##   Teacher                       0  38        0  48
```

```
library(lattice)
with(subset(ecommerce_df, profile %in% c("Parent", "Teacher") &
productKnewWhatWanted %in% c("No", "Yes")), histogram( ~
productKnewWhatWanted | profile))
```



#Answer 2 Yes, I should limit the observation product knowledge of “Yes” or “No”, because propotion of “Somewhat” in the group is almost same, and hence, doesn’t change the ratio/propotion of “Yes” or “No” in the groups for our study

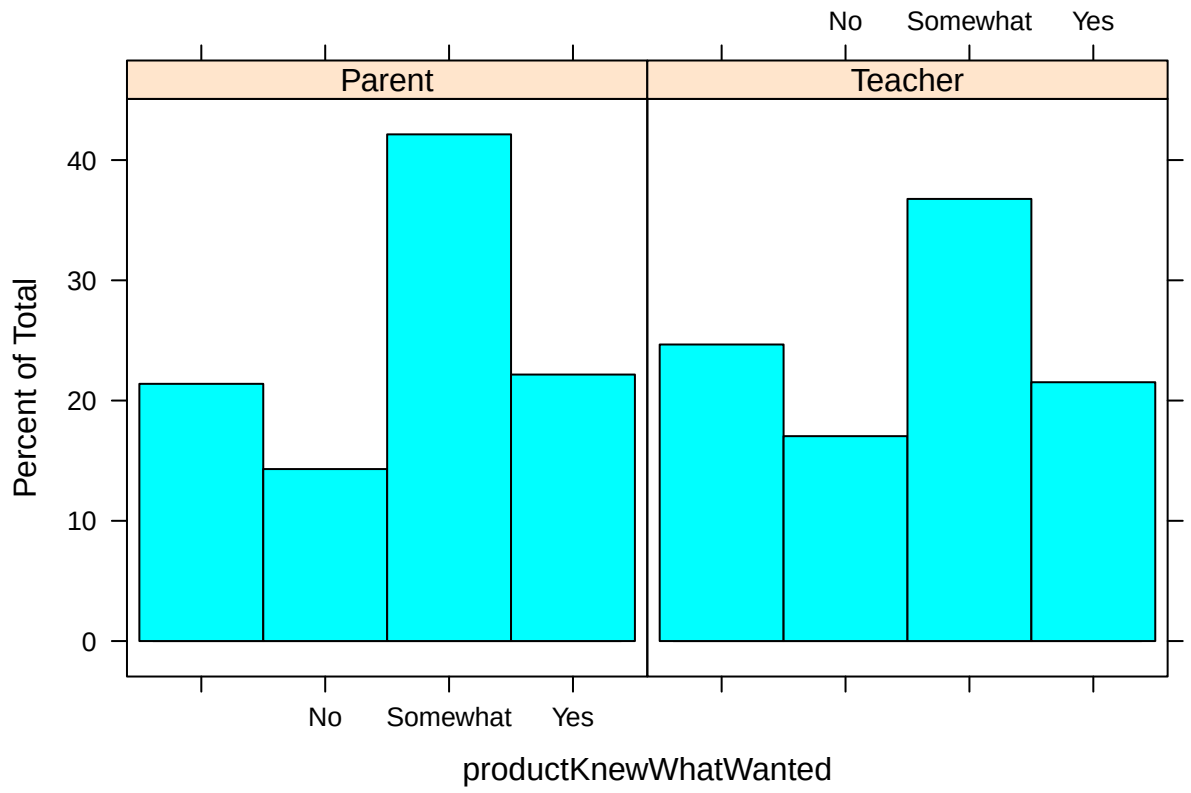
```r
with(subset(ecommerce_df, profile %in% c("Parent", "Teacher") &
productKnewWhatWanted %in% c("No", "Yes")), prop.table(table(profile,
productKnewWhatWanted),margin=1))
```

```
##                            productKnewWhatWanted
## profile                                       No  Somewhat       Yes
##   0
##   Friend/family friend
##   Health Professional
##   Other
##   Parent                    0.0000000 0.3922261 0.0000000 0.6077739
##   Person with [condition A]
##   Relative
##   Teacher                   0.0000000 0.4418605 0.0000000 0.5581395
```

```r
with(subset(ecommerce_df,
            profile %in% c("Parent", "Teacher")),
     prop.table(table(profile, productKnewWhatWanted), margin=1))
```

```
##                            productKnewWhatWanted
## profile                                       No  Somewhat       Yes
##   0
##   Friend/family friend
##   Health Professional
##   Other
##   Parent                    0.2139175 0.1430412 0.4213918 0.2216495
##   Person with [condition A]
##   Relative
##   Teacher                   0.2466368 0.1704036 0.3677130 0.2152466
```

```r
with(subset(ecommerce_df,
            profile %in% c("Parent", "Teacher") ),
     histogram( ~ productKnewWhatWanted | profile))
```



#Answer 3 No, the difference in prior product knowledge in groups Parent and Teachers is not significant.

```
prod.table <- with(subset(ecommerce_df,
            profile %in% c("Parent", "Teacher") & productKnewWhatWanted %in% c("No", "Yes")),
     (table(profile, productKnewWhatWanted)))

prod.table[c(5, 8), c(2, 4)]
```

```
##          productKnewWhatWanted
## profile    No Yes
##   Parent  111 172
##   Teacher  38  48
```

```
prop.table(prod.table[c(5, 8), c(2, 4)], margin=1)
```

```
##          productKnewWhatWanted
## profile          No       Yes
##   Parent  0.3922261 0.6077739
##   Teacher 0.4418605 0.5581395
```

```
chisq.test(prod.table[c(5, 8), c(2, 4)])
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
```

```
## data:  prod.table[c(5, 8), c(2, 4)]
## X-squared = 0.48452, df = 1, p-value = 0.4864
```

#Answer 4 Propotion of Teacher with prior product knowledge is 55.8% and propotion of Parent with prior knowledge is 60.8%

```
prop.table(prod.table[c(5, 8), c(2, 4)], margin=1)
```

```
##          productKnewWhatWanted
## profile          No       Yes
##   Parent  0.3922261 0.6077739
##   Teacher 0.4418605 0.5581395
```

#Answer 5 No the percentage is not statistically significant, Based on our data the 95% confidence interval is 45%-67%, which includes the null hypothesis value of 60.7%.

```
binom.test(prod.table[8, 4],
           prod.table[8, 4] + prod.table[8, 2],
           p = prop.table(prod.table, margin=1)[5, 4] )
```

```
##
##  Exact binomial test
##
## data:  prod.table[8, 4] and prod.table[8, 4] + prod.table[8, 2]
## number of successes = 48, number of trials = 86, p-value = 0.3773
## alternative hypothesis: true probability of success is not equal to 0.6077739
## 95 percent confidence interval:
##  0.4469536 0.6652055
## sample estimates:
## probability of success
##              0.5581395
```

#Answer 6 PageViews is significantly different in the group Parent and Teacher, Mean of Parent is higher i.e. 6.28 > Mean of Teacher i.e. 5.46. The 95% confidence interval for difference is 0.28 - 1.36

```
pageViewInt <- rep(NA, length(ecommerce_df$behavPageviews))
pageViewInt[ecommerce_df$behavPageviews=="0"]      <- 0
pageViewInt[ecommerce_df$behavPageviews=="1"]      <- 1
pageViewInt[ecommerce_df$behavPageviews=="2 to 3"] <- 2
pageViewInt[ecommerce_df$behavPageviews=="4 to 6"] <- 4
pageViewInt[ecommerce_df$behavPageviews=="7 to 9"] <- 7
pageViewInt[ecommerce_df$behavPageviews=="10+"]    <- 10
ecommerce_df$pageViewInt <- pageViewInt
rm(pageViewInt)
with(subset(ecommerce_df,
            profile %in% c("Parent", "Teacher")),
     t.test(pageViewInt ~ profile))
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  pageViewInt by profile
## t = 2.9841, df = 361.1, p-value = 0.003038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.279285 1.358804
## sample estimates:
##  mean in group Parent mean in group Teacher
##              6.280928              5.461883
```

#Answer 7 The pageViewInt varies significantly by profiles, because F value is less than 0.05 and also the mean values are different

```
prof.aov <- aov(pageViewInt ~ -1+profile, data=ecommerce_df)
anova(prof.aov)
```

```
## Analysis of Variance Table
##
## Response: pageViewInt
##             Df Sum Sq Mean Sq F value    Pr(>F)
## profile      8  53367  6670.9   513.9 < 2.2e-16 ***
## Residuals 1585  20575    13.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

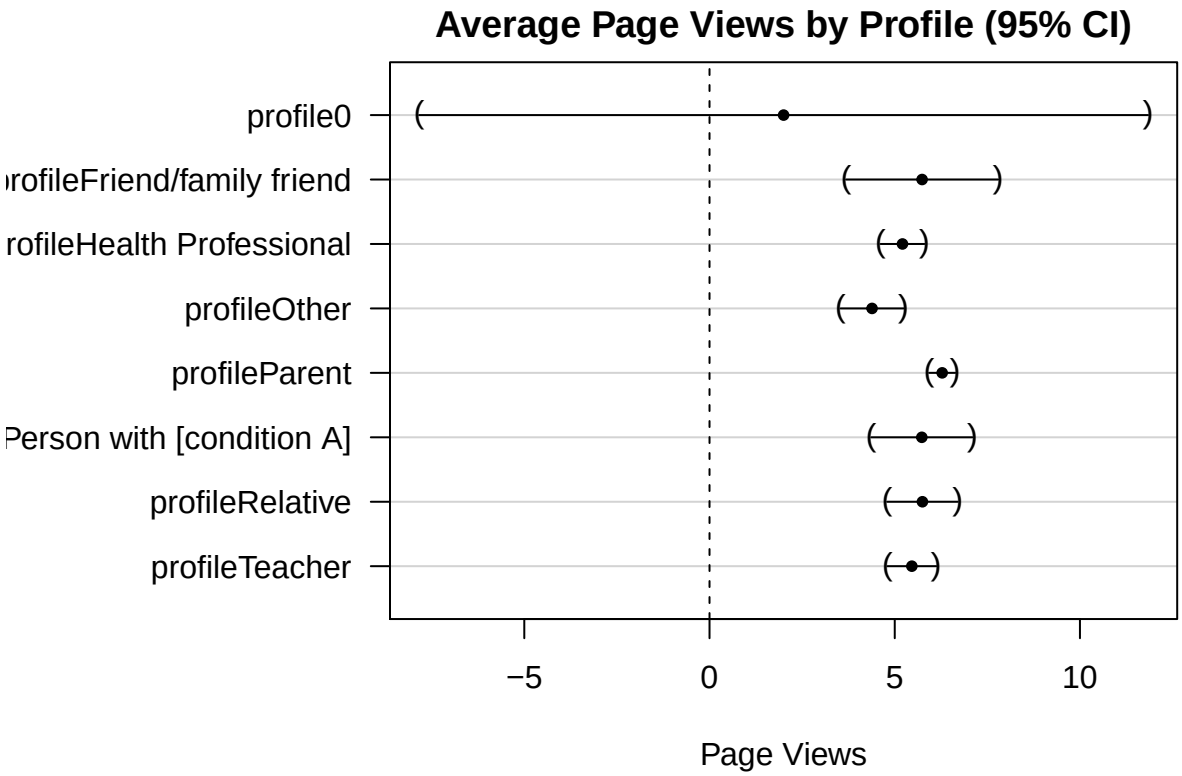
```
library(multcomp)
```

```
## Warning: package 'multcomp' was built under R version 3.6.3

## Loading required package: mvtnorm

## Loading required package: survival

## Loading required package: TH.data

## Warning: package 'TH.data' was built under R version 3.6.3

## Loading required package: MASS

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##     geyser
```

```
glht(prof.aov)
```

```
##
##   General Linear Hypotheses
##
## Linear Hypotheses:
##                                       Estimate
## profile0 == 0                            2.000
## profileFriend/family friend == 0         5.739
## profileHealth Professional == 0          5.209
## profileOther == 0                        4.388
## profileParent == 0                       6.281
## profilePerson with [condition A] == 0    5.731
## profileRelative == 0                     5.748
## profileTeacher == 0                      5.462
```

```
par(mar=c(6,10,2,2)) # adjusts margins to preserve axis labels
plot(glht(prof.aov), xlab="Page Views", main="Average Page Views by Profile (95% CI)")
```



#Answer 8 The dot shows the mean for each segment i.e. 5.462 and 6.281, and bars reflect the confidence interval. Confidence intervals for the mean pageViews of each segment are shown. It is clear that the average PageViews of Parents is substantially greater than the Teacher. No, the mean value is not different than the previous question in the above analysis because the mean within the group is not changing, keeping the confidence interval same

```
prof.aov2 <- aov(pageViewInt ~  -1+profile, data=subset(ecommerce_df,
                                                        profile %in% c("Parent", "Teacher")))
anova(prof.aov2)
```

```
## Analysis of Variance Table
##
## Response: pageViewInt
##            Df Sum Sq Mean Sq F value    Pr(>F)
## profile     2  37266 18632.9  1418.7 < 2.2e-16 ***
## Residuals 997  13094    13.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
glht(prof.aov2)
```

```
##
##   General Linear Hypotheses
##
## Linear Hypotheses:
##                       Estimate
## profileParent == 0     6.281
## profileTeacher == 0    5.462
```

```
par(mar=c(6,10,2,2)) # adjusts margins to preserve axis labels
plot(glht(prof.aov2), xlab="Page Views", main="Average Page Views by Profile, Teachers/Parents (95% CI)")
```