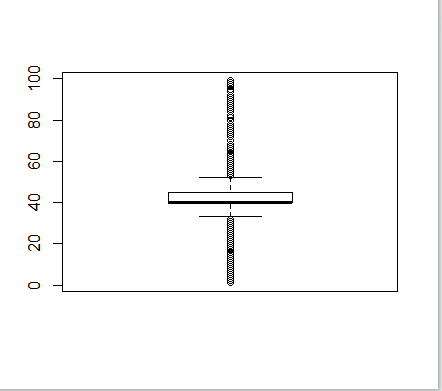
This question is on the basis of the  [adult.data.simplified.csv](https://blackboard.valpo.edu/bbcswebdav/pid-440738-dt-content-rid-2539362_1/xid-2539362_1) 

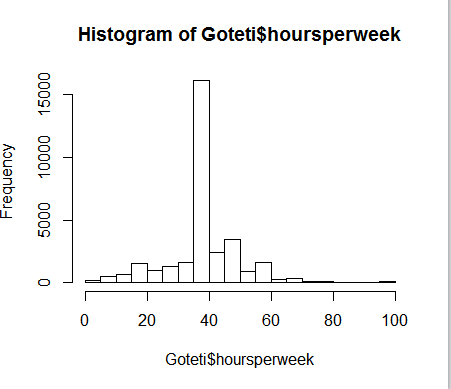
Using R for questions 1-3 and MS Excel for question 4, please answer the questions below.  Be sure to include your R code and a screenshot of your MS Excel solution.

**ENTER YOUR ANSWERS IN THE FIELD BELOW.  FOR THE SCREENSHOTS AND GRAPHICS, PLEASE ATTACH A SEPARATE FILE.**

1. How are the attribute values distributed?  Build a boxplot and a histogram for each.  You will have to transform the nominal attribute so you can plot its values.

Attributes values are being ranging from 1 to 99 in hours per week attribute, which is said to be considered as minimum value as 1 and maximum is 99 and remaining values are been organized in this range, median as 40.00 and mean as 40.44 where hours per week are increasing however the sex is internally related to hours per week, and it is increasing per quartile as well. Taking the nominal value as relationship in which number of husband working is more than number of wifes, and where sex is getting internally related to the people working.  
   
boxplot(Goteti$hoursperweek)

  
Fig: blogspot for $hoursperweek

  
Fig: histogram of hours per week

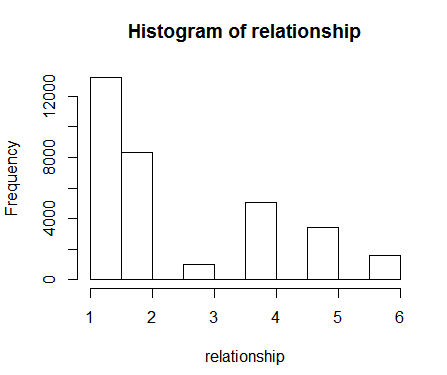
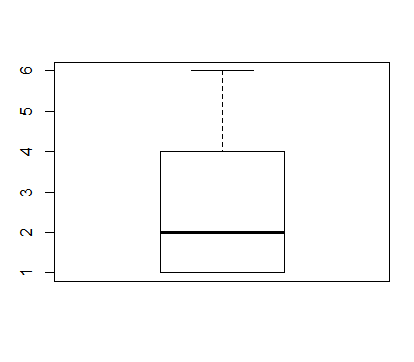
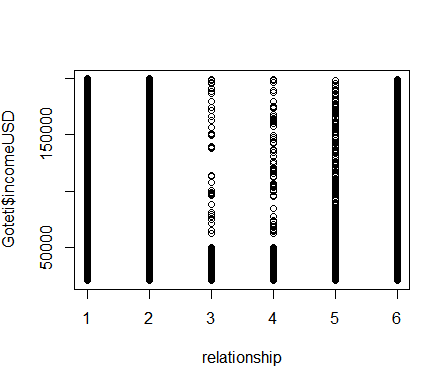
Goteti$relationship=as.numeric(Goteti$relationship)

is.numeric(Goteti$relationship)

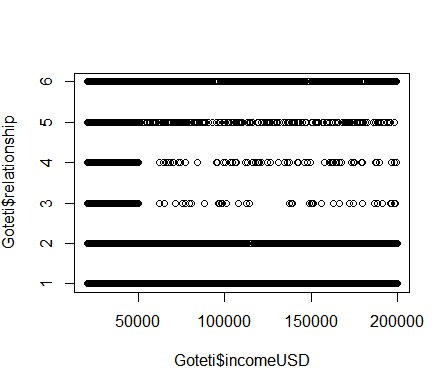
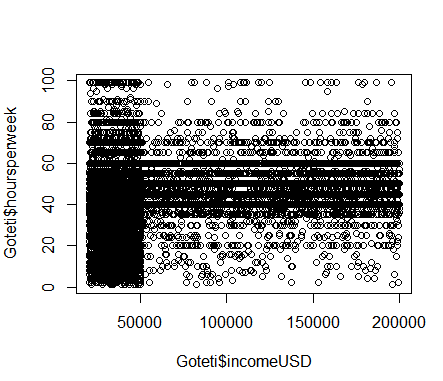
hist(relationship)

plot(relationship,Goteti$incomeUSD)

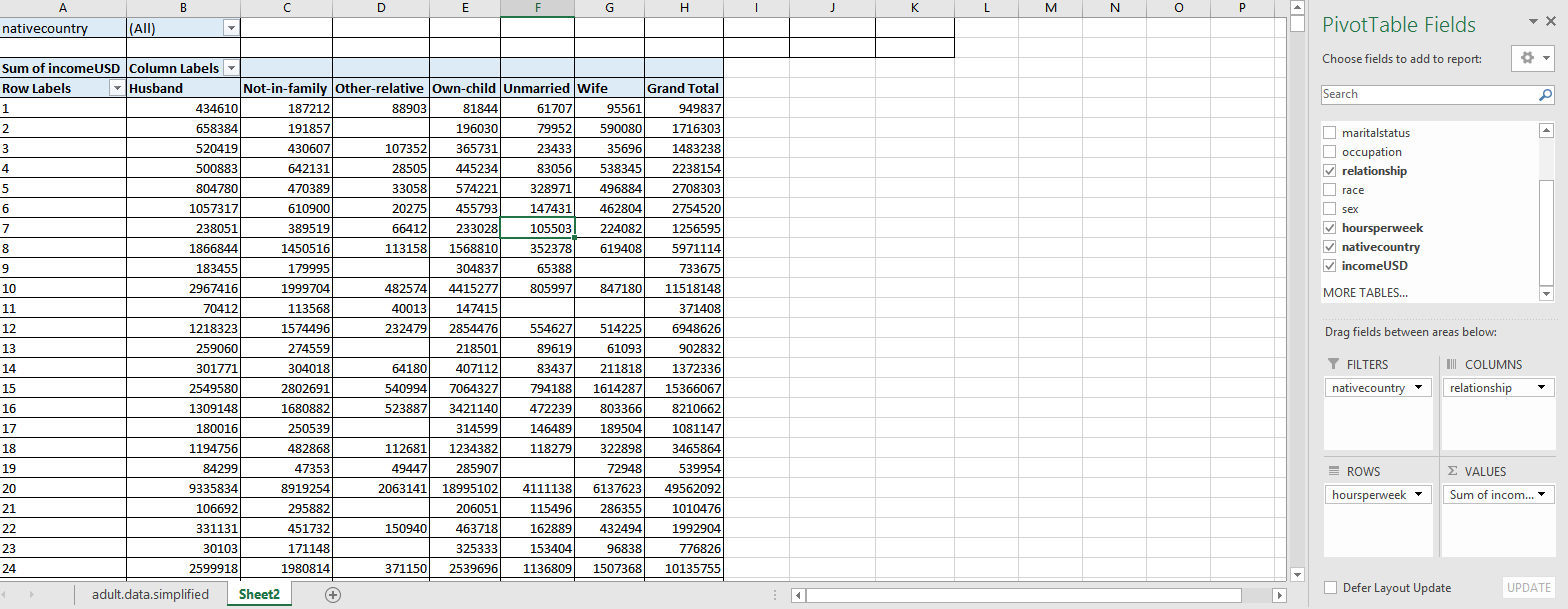
boxplot(relationship)

  
  
  
  
 Fig: Boxplot of relation  
  
  
Fig: Scatter plot of relationship and IncomeUSD   
  
  
  
  
2.How does EACH OF the two attributes relate to the “income” attribute?  Build a scatter diagram for EACH relationship and determine which attribute is more closely related..  
  
  
  
plot(Goteti$incomeUSD,Goteti$relationship)

plot(Goteti$incomeUSD,Goteti$hoursperweek)  
  
  
Relationship and hours perweek for people who are not in family and for wifes, and where it is internally related to each other to have more income earned than other values in this attributes. As the hours per week is high other memebers in the family , is earning more than the husband of the family.

3.In MS Excel, build a Pivot chart showing aggregated data for your two attributes and income by country.  What can we learn from this roll-up?

There are some couple of inferences from this, and people who does not belong to the family, and Varied country are earning more than the husband, and wifes and children.  
  
  
  
  
  
  
  
Source Code:   
  
Goteti<- read.csv("./adult.data.simplified.csv")

attributes(Goteti)

str(Goteti)

summary(Goteti$relationship)

boxplot(Goteti$hoursperweek)

sd(Goteti$hoursperweek)

hist(Goteti$hoursperweek)

boxplot(Goteti$relationship)

hist(Goteti$relationship)

summary(Goteti$relationship)

plot(Goteti$hoursperweek)

summary(Goteti$hoursperweek)

range(Goteti$hoursperweek)

plot(Goteti$incomeUSD,Goteti$relationship)

plot(Goteti$incomeUSD,Goteti$hoursperweek)

Goteti$race=as.numeric(Goteti$relationship)

is.numeric(Goteti$relationship)

hist(Goteti$relationship)

plot(Goteti$relationship,Goteti$incomeUSD)

boxplot(relationship)