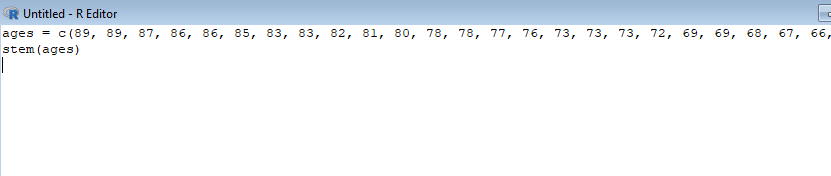
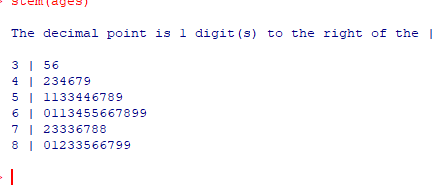
**Homework 1 (Due: Friday, Feb 08)**  
  
**Guideline:** **Please include the screen shot of R-syntax and the results as presented in Lab-3 for each part of the question except 3(a) for which you just provide R-syntax.**-------------------------------------------------------------------------------------------------------------------------------

1. The following represents the ages of the 50 richest people in the world in 2009.  
  
89, 89, 87, 86, 86, 85, 83, 83, 82, 81, 80, 78, 78, 77, 76, 73, 73, 73, 72, 69, 69, 68, 67, 66, 66, 65, 65, 64, 63, 61, 61, 60, 59, 58, 57, 56, 54, 54, 53, 53, 51, 51, 49, 47, 46, 44, 43, 42, 36, 35

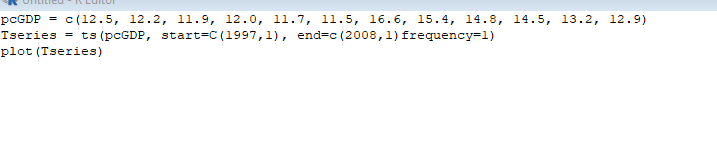
Use a stem-and-leaf plot to organize the ages of the 50 richest people data set listed above. What can you conclude?

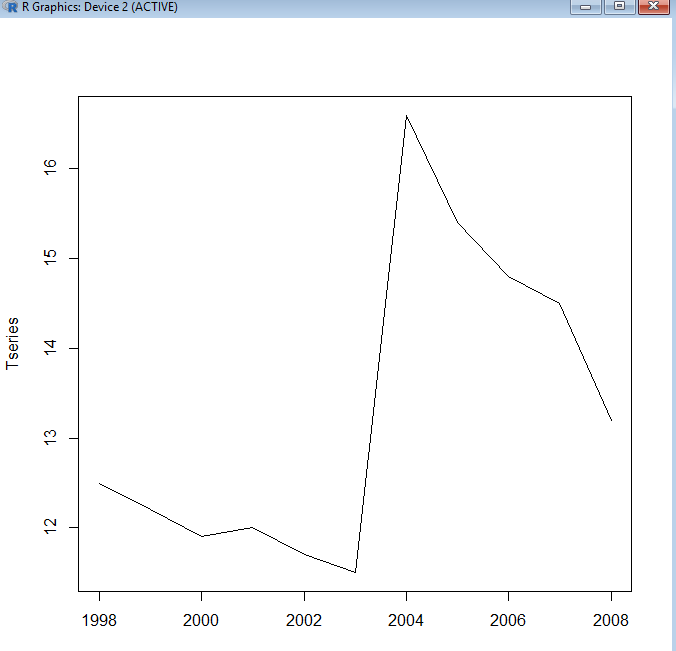


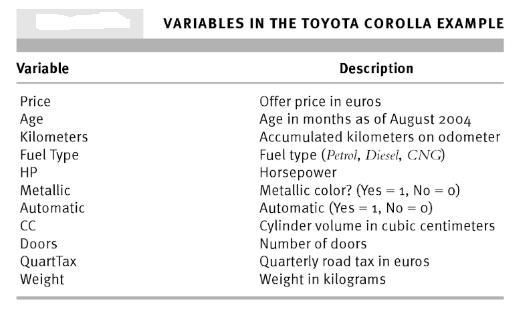


You can conclude, that most of the top 50 richest people are in their 60s

2. Use a time series chart to display the data. The data represent the percentages of the U.S. gross domestic product (GDP) that come from the manufacturing sector.

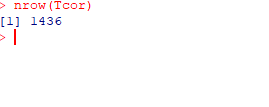
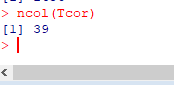
1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008   
12.5% 12.2% 11.9% 12.0% 11.7% 11.5% 16.6% 15.4% 14.8% 14.5% 13.2% 12.9%   




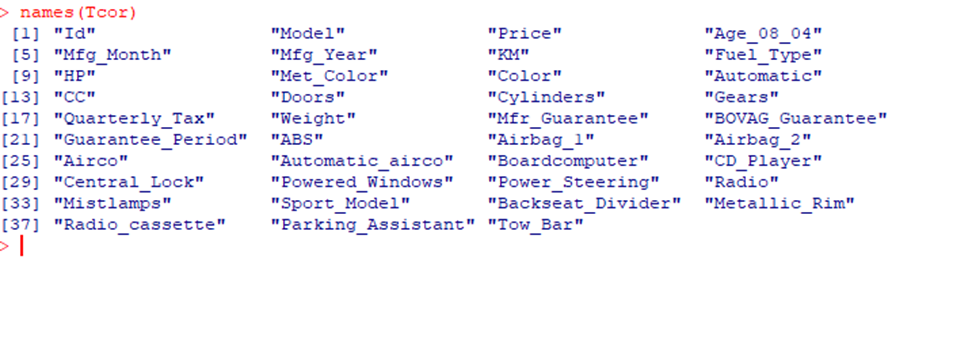
3. A large Toyota car dealership offers purchasers of new Toyota cars the option to buy their used car as part of a trade-in. In particular, a new promotion promises to pay high prices for used Toyota Corolla cars for purchasers of a new car. The dealer then sells the used cars for a small profit. To ensure a reasonable profit, the dealer needs to be able to predict the price that the dealership will get for the used cars. For that reason, data were collected on all previous sales of used Toyota Corollas at the dealership. The data include the sales price and other information on the car, such as its age, mileage, fuel type, engine size, etc. The description of the part of the variables is as follows  
  
  
The data file is **ToyotaCorolla.csv**. In this data set I would like you to perform the following tasks:  
  
a. Read the data file in R to create a SAS data set. Name it **TCor**.

Tcor = read.csv("C:/Users/russ\_/Documents/Git/SCHOOL/SP2019/MATH315/Homework/Homework1/ToyotaCorolla.csv", header=T, sep=",")

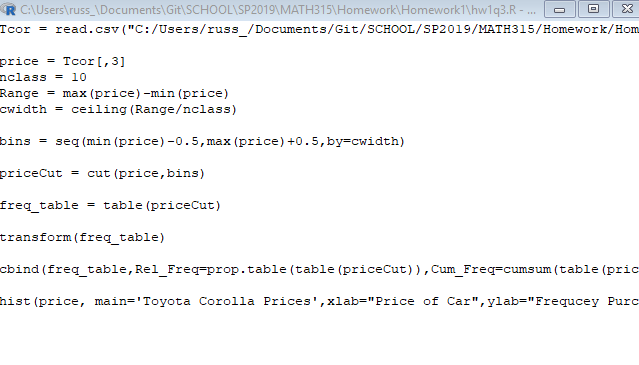
b. In **TCor** data, determine the number of variables and the number of data rows (also called number of observations or data records)

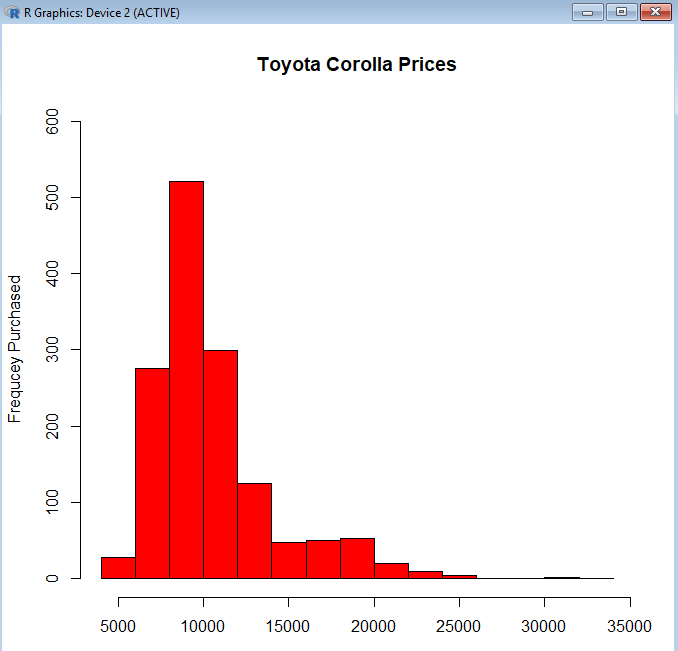
.

c. List all the variable names.

  
  
d. Identify the type of variables (qualitative vs quantitative). List all the variables of each type.

e. Construct a histogram of the **Price** variable. Make sure to include the title, x-axis label and y-axis label and appropriate limits of y-axis. Describe the shape of the data distribution (symmetric, skewed right/left).

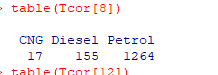




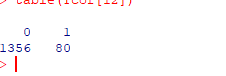
The price is definitely skewed to one side with most people paying around 10000 and the trend tapering as the price of the car gets higher.

f. Construct relative frequency distributions of variables: **Fuel\_Type and Automatic**.

Gas Type:



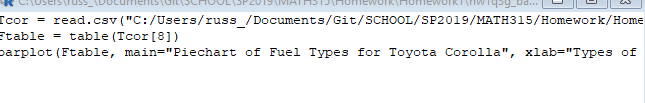
Manual/Automatic

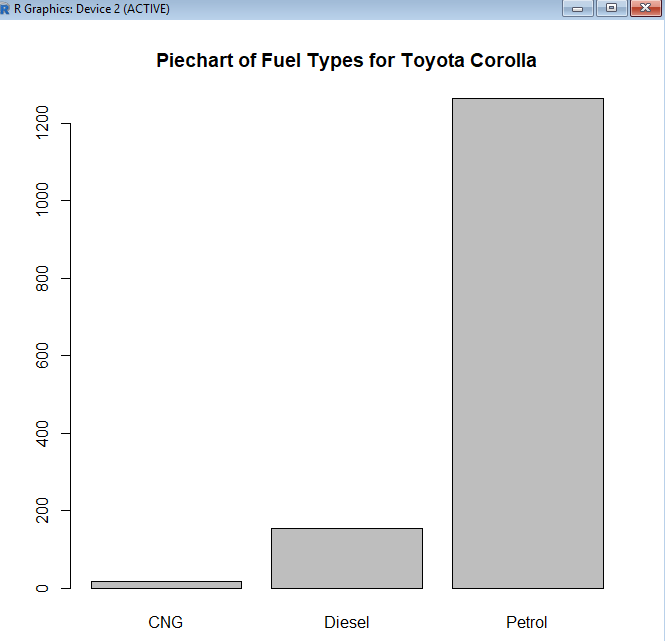


g. Construct a barplot and a pie chart of **Fuel\_Type**. Interpret the plots.

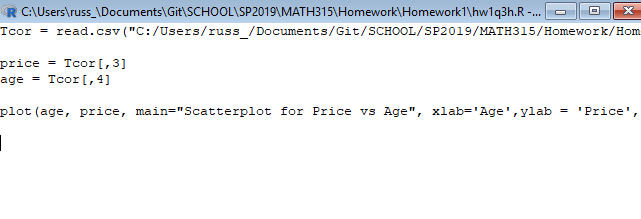
Pie:

Barplot:





h. Construct a scatter plot of Price and Age\_08\_04. Describe the relationship between these two variables.





This tells us that younger people are willing to spend more on a Toyota Corolla, with people older than 60 only willing to spend around 10000 dollars on this car.